

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 21, 2004, 05:50:54 ; Search time 535 Seconds
(without alignments)
9890.498 Million cell updates/sec

Title: US-10-617-962-2
Perfect score: 1008
Sequence: 1 atggttatacaattacacac.....ctcatataaataatataga 1008

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4134886 seqs, 2624710521 residues

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_23Sep04:*

1: geneseqn1980s:*\n2: geneseqn1980s:*\n3: geneseqn2000s:*\n4: geneseqn2001as:*\n5: geneseqn2001bs:*\n6: geneseqn2002as:*\n7: geneseqn2002bs:*\n8: geneseqn2003as:*\n9: geneseqn2003bs:*\n10: geneseqn2003cs:*\n11: geneseqn2003ds:*\n12: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1008	100.0	1008	2	AAX07350 Photorhab
2	1008	100.0	1388	2	AAX07352 Photorhab
3	992	98.4	110000	10	ACF68075 Photorhab
4	992	98.4	110000	10	ACF67367 07
5	992	98.4	110000	10	ACF65384_1
6	663.8	65.9	11007	2	AAX07349 Xenorhabd
7	663.8	65.9	1205	2	AAX07351 Xenorhabd
8	628.2	62.3	1272	2	AA080839 Xenorhabd
9	52.6	5.2	3927	3	AA070101 Plasmodi
10	51	5.1	8056	8	AB210246 Haematopo
11	50.8	5.0	4590	1	AA060472 Sequence
12	50.6	5.0	188971	12	ADL08108 Human gen
13	50.4	5.0	3738	3	AA070178 Plasmodi
14	49.8	4.9	11394	6	ABK28221 DNA trans
15	49.2	4.9	5940	3	AA070105 Plasmodi
16	48.8	4.8	1866	2	AA099653 Nucleic a
17	48.2	4.8	2767	10	ADG42381 DNA encod
18	47.8	4.7	729	6	ABN91698 Staphyloc
19	47.6	4.7	1953	12	AB019011 DNA encod
20	47.6	4.7	16217	6	ABJ32624 Human imm
21	47.4	4.7	717	8	ACF73065 Staphyloc

C	22	47.4	4.7	41708	3	AA068247 Bacteriop
C	23	47.4	4.7	41708	4	AA068106 Complete
C	24	47.4	4.7	158001	12	ADL17884 Human pho
C	25	47.2	4.7	700	4	AA093026 Human inf
C	26	47	4.7	549	3	AA093384 Cat flea
C	27	47	4.7	5340	5	AA062829 P. falcip
C	28	47	4.7	11052	6	ABK39986 Human che
C	29	46.4	4.6	5409	3	AA070151 Plasmodi
C	30	46.4	4.6	8056	8	AB210100 Haematopo
C	31	46	4.6	32392	6	ABL56203 AmsEPV gen
C	32	45.6	4.5	18283	6	ABL70502 Chemical1
C	33	45.6	4.5	18283	6	AA061363 Human 3'
C	34	45.4	4.5	605	2	AA031530 Human
C	35	45.4	4.5	9824	10	ADP42684 Plasmid v
C	36	45.2	4.5	4985	6	AB075107 Anopheles
C	37	45.2	4.5	4985	10	ACF79720 Mosquito
C	38	45.2	4.5	13377	6	AA046476 Tumour su
C	39	45.2	4.5	13377	6	ABL33463 Human imm
C	40	45	4.5	583	8	ADA72369 Rice gene
C	41	45	4.5	6103	6	ABL33691 Human imm
C	42	44.6	4.4	1488	4	AA052007 Staphyloc
C	43	44.6	4.4	1488	8	ACF74231 Staphyloc
C	44	44.6	4.4	1491	4	AA054922 Staphyloc
C	45	44.6	4.4	1491	8	ACA20304 Prokaryot

ALIGNMENTS

RESULT 1
AAX07350 standard; DNA; 1008 BP.

21-MAY-1999 (first entry)

Photorhabdus luminescens insecticidal toxin gene P1V16tox1.

Toxin; P1V16tox1; biological control; insecticide; ds.

Photorhabdus luminescens.

MO9903328-A1.

28-JAN-1999.

17-JUL-1998; 98WO-AU000562.

17-JUL-1997; 97AU-0008088.

(CSTR) COMMONWEALTH SCI & IND RES ORG.

Eact PD;

WPI; 1999-131733/11.

P-PSDB; AAM97811.

New insecticidal toxin genes - extracted from Xenorhabdus nematophilus

A24 and Photorhabdus luminescens.

Claim 1; Page 27-28; 48pp; English.

This is the coding region of the insecticidally-active toxin gene, termed P1V16tox1, of Photorhabdus luminescens strain V16/1. It codes for a 335-amino acid P1V16tox1 protein (see AAM97811). The gene was isolated from a genomic DNA cosmid library using the Xenorhabdus nematophilus A24 tox24 gene (see AAX07349) as probe. A 1.39 kb HindIII/SmaI fragment (see AAX07352) of the isolated gene included the 1008 bp coding region. The invention relates to the identification and isolation of polynucleotide molecules encoding a new class of protein insecticidal toxins which are produced by bacteria of the genera Xenorhabdus and Photorhabdus. The polynucleotide molecules may be incorporated e.g. into insect-specific

CC viruses (e.g. entomopox and nuclear polyhedrosis viruses), bacteria (e.g. CC Gracilicutes, Firmicutes, Tenericutes and Mendosicutes), protozoa, yeast CC and plants (particularly cereals such as wheat and barley, vegetables CC such as tomato and potato, fruit trees such as citrus and apple, tobacco CC and cotton) for control of pest insects

XX Sequence 1008 BP; 376 A; 146 C; 192 G; 294 T; 0 U; 0 Other;

Query Match 100.0%; Score 1008; DB 2; Length 1008;
Best Local Similarity 100.0%; Pred. No. 2e-210;
Matches 1008; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGTTATACATTAACACCTGATGATAGAGTATATCCACCCGTTGAAAAACAATA 60
DB 1 ATGGTTATACATTAACACCTGATGATAGAGTATATCCACCCGTTGAAAAACAATA 60
QY 61 GAGAGAGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 120
DB 61 GAGAGAGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 120
QY 121 TATGATATGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 180
DB 121 TATGATATGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 180
QY 181 ATTCATATGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 240
DB 181 ATTCATATGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 240
QY 241 ATTTATGAGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300
DB 241 ATTTATGAGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300
QY 301 ATTTATGAGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360
DB 301 ATTTATGAGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360
QY 361 CCTCAAAACATTTGGAAATGTTCTGAGCTTGAATAAACAATGATGATGATGATGATGAT 420
DB 361 CCTCAAAACATTTGGAAATGTTCTGAGCTTGAATAAACAATGATGATGATGATGATGAT 420
QY 421 GACGATATATTTTGGCACTATATTTTCTGATGAGAAATTCACATGAGAGAAAT 480
DB 421 GACGATATATTTTGGCACTATATTTTCTGATGAGAAATTCACATGAGAGAAAT 480
QY 481 CAACATCAAAATGCGCAAGATTTTAAATTAATGATGATGATGATGATGATGATGATGAT 540
DB 481 CAACATCAAAATGCGCAAGATTTTAAATTAATGATGATGATGATGATGATGATGATGAT 540
QY 541 GTAACCTTCACTGGGAGAGATTTTCAAAAACCTTTTACATGATGATGATGATGATGATGAT 600
DB 541 GTAACCTTCACTGGGAGAGATTTTCAAAAACCTTTTACATGATGATGATGATGATGATGAT 600
QY 601 TCATTAGAGATTTATTTGAGAGAAAAAATCTTTCAAACTTTCTTTCGACACCGAG 660
DB 601 TCATTAGAGATTTATTTGAGAGAAAAAATCTTTCAAACTTTCTTTCGACACCGAG 660
QY 661 AGATTACCTGATGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 720
DB 661 AGATTACCTGATGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 720
QY 721 GTGAGTTTAAAGAACTTAAATAAACAATCTAGAGATGATGATGATGATGATGATGATGATGAT 780
DB 721 GTGAGTTTAAAGAACTTAAATAAACAATCTAGAGATGATGATGATGATGATGATGATGATGAT 780
QY 781 GGTGAGAAACAAAGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 840
DB 781 GGTGAGAAACAAAGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 840
QY 841 ACAGAGAGAGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 900
DB 841 ACAGAGAGAGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 900
QY 901 TTATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 960

DB 901 TTATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 960
QY 961 AATCAATGAGGTTTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1008
DB 961 AATCAATGAGGTTTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1008

RESULT 2
AXX07352
ID AAX07352 standard; DNA; 1388 BP.

XX AAX07352;
XX 21-MAY-1999 (first entry)
XX Photorhabdus luminescens insecticidal toxin gene pVI6tox1.
XX Toxin; pVI6tox1; biological control; insecticide; db.
XX Photorhabdus luminescens.
XX OS
XX Key Location/Qualifiers
XX CDS 172..1178
XX FT /*tag= a
XX FT /note= "the coding region is specifically claimed in
XX Claim 1"

PN WO9903328-A1.
XX 28-JAN-1999.
XX 17-JUL-1998; 98WO-AU000562.
XX 17-JUL-1997; 97AU-00008088.
XX (CSIR) COMMONWEALTH SCI & IND RES ORG.
XX East PD;
XX WPI, 1999-131733/11.
XX P-PSDB; AAX97811.

PT New insecticidal toxin genes - extracted from Xenorhabdus nematophilus
XX A24 and Photorhabdus luminescens.
XX Example 1; Page 32; 48pp; English.
XX This is the nucleotide sequence of the insecticidally-active toxin gene,
XX termed pVI6tox1, of Photorhabdus luminescens strain V16/1. It includes a
XX coding region, which is claimed (see AAX07350), that codes for a 335-
XX amino acid pVI6tox1 protein (see AAX97811). The gene was isolated from a
XX genomic DNA cosmid library using the Xenorhabdus nematophilus A24 toxB4
XX gene (see AAX07349) as probe. The invention relates to the identification
XX and isolation of polynucleotide molecules encoding a new class of protein
XX insecticidal toxins which are produced by bacteria of the genera
XX Xenorhabdus and Photorhabdus. The polynucleotide molecules may be
XX incorporated e.g. into insect-specific viruses (e.g. entomopox and
XX nuclear polyhedrosis viruses), bacteria (e.g. Gracilicutes, Firmicutes,
XX Tenericutes and Mendosicutes), protozoa, yeast and plants (particularly
XX cereals such as wheat and barley, vegetables such as tomato and potato,
XX fruit trees such as citrus and apple, tobacco and cotton) for control of
XX pest insects

XX Sequence 1388 BP; 501 A; 213 C; 261 G; 413 T; 0 U; 0 Other;

Query Match 100.0%; Score 1008; DB 2; Length 1388;
Best Local Similarity 100.0%; Pred. No. 2.1e-210;
Matches 1008; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGTTATACATTAACACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 60
DB 172 ATGGTTATACATTAACACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 231

```

QY 61 GCAGAGATATAGTACGTATCTAACTTTAGCAACAGATGAGGCTCATACGATCA 120
DB 232 GCAGAGATATAGTACGTATCTAACTTTAGCAACAGATGAGGCTCATACGATCA 291
QY 121 TATGGAATTGAAATCGAGCTAGAGAAATTAATATTAGCTTACGCTTGGCTGAAGTGT 180
DB 292 TATGGAATTGAAATCGAGCTAGAGAAATTAATATTAGCTTACGCTTGGCTGAAGTGT 351
QY 181 ATTCAATATGATCTAAACCTTCCTGATGACTATTATAGAAATTAAGAGACTGCTGAGGA 240
DB 352 ATTCAATATGATCTAAACCTTCCTGATGACTATTATAGAAATTAAGAGACTGCTGAGGA 411
QY 241 ATTATCAAGAAATATATGCTTAATCTTCATCTGCACTATTAGAGAAATGAGATCA 300
DB 412 ATTATCAAGAAATATATGCTTAATCTTCATCTGCACTATTAGAGAAATGAGATCA 471
QY 301 ATTTCTAAAGATATGCAAAATGCTTTTATAGAAATGAACTGATTTTGAAGTCAATAT 360
DB 472 ATTTCTAAAGATATGCAAAATGCTTTTATAGAAATGAACTGATTTTGAAGTCAATAT 531
QY 361 CCTCAAAACATTTGGATGTTCTGAGCTTGAAATTAACCAATGAGTCTTATTCAGAT 420
DB 532 CCTCAAAACATTTGGATGTTCTGAGCTTGAAATTAACCAATGAGTCTTATTCAGAT 591
QY 421 GACGATTAATATATGCACTATATTTTCTCTGACAGGAAATCCACTGAGGAAAT 480
DB 592 GACGATTAATATATGCACTATATTTTCTCTGACAGGAAATCCACTGAGGAAAT 651
QY 481 CAACATCAAAATGCCGCAAGATTTTAAATTAATTTGATTTCTTATACCTTATCTGCT 540
DB 652 CAACATCAAAATGCCGCAAGATTTTAAATTAATTTGATTTCTTATACCTTATCTGCT 711
QY 541 GTAACCTCACTGGAGAGAGAGATTTTCAAAAACTTTTCAATGATTAAGAGCTTAA 600
DB 712 GTAACCTCACTGGAGAGAGATTTTCAAAAACTTTTCAATGATTAAGAGCTTAA 771
QY 601 TCATTAAGAAATTAATGAGAGAAATTAATCTTAACTTAACTTTTCCGACCCGAG 660
DB 772 TCATTAAGAAATTAATGAGAGAAATTAATCTTAACTTAACTTTTCCGACCCGAG 831
QY 661 AGATTACCTGATGAGAGATAGTATTATTTGGCTGAGCAACAGAGCCCTTAATGAGA 720
DB 832 AGATTACCTGATGAGAGATAGTATTATTTGGCTGAGCAACAGAGCCCTTAATGAGA 891
QY 721 GTGAGTTTAAAGAACTTAAATAAATAAATACTAGGAATGATTTTCTAATATGAGAGG 780
DB 892 GTGAGTTTAAAGAACTTAAATAAATAAATACTAGGAATGATTTTCTAATATGAGAGG 951
QY 781 GCTGCAAAACAAAGATATAGTCTTATTAATTAAGAGCAAAAGGTTACGCTCCAG 840
DB 952 GCTGCAAAACAAAGATATAGTCTTATTAATTAAGAGCAAAAGGTTACGCTCCAG 1011
QY 841 ACAGAGAGGAAAGATATGCTACAGCCAGTGCATTAACCTTGAATAATTCGCAATAT 900
DB 1012 ACAGAGAGGAAAGATATGCTACAGCCAGTGCATTAACCTTGAATAATTCGCAATAT 1071
QY 901 TTATATAGTGTAGGCTTAAGCCAAAGAGAGGTAACCTTTACTCAAAATGATATGAC 960
DB 1072 TTATATAGTGTAGGCTTAAGCCAAAGAGAGGTAACCTTTACTCAAAATGATATGAC 1131
QY 961 AATAAATAGAGGCTTCAATGCTTGAAGCTCATTAATAAATAATATATGA 1008
DB 1132 AATAAATAGAGGCTTCAATGCTTGAAGCTCATTAATAAATAATATATGA 1179

```

RESULT 3
ACF68075 standard; DNA; 1008 BP.
AC ACF68075;
XX
DT 20-NOV-2003 (first entry)

```

XX XX Photorhabdus luminescens nucleotide sequence #6542.
DE Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;
XX detection; food; gene expression; plant; animal; microorganism; toxin;
KW antibiotic; biopesticide; virulence factor; disease model; plague;
KW whooping cough; gene; ds.
XX OS Photorhabdus luminescens.
XX WO200294867-A2.
XX 28-NOV-2002.
XX 07-FEB-2002; 2002WO-1B003040.
XX 07-FEB-2001; 2001FR-00001659.
XX (INSP ) INST PASTEUR.
XX (CNRS ) CNRS CENT NAT RECH SCI.
XX Duchaud E, Taourit S, Glaeser P, Frangeul L, Kunst F, Danchin A,
XX Buchrieser C;
XX WPI; 2003-148459/14.
XX
XX Genomic sequence of Photorhabdus luminescens and encoded polypeptides,
XX useful e.g. as therapeutic antimicrobials and agricultural pesticides.
XX
XX Claim 2; SEQ ID NO 6542; 1205bp; French.
XX
XX The invention relates to the isolation of genes and their encoded
XX proteins from Photorhabdus luminescens. The isolated sequences are
XX sources of probes and primers for detecting the genome of P. luminescens
XX and related species; to study polymorphisms; for gene analysis and for
XX detection/amplification of the genes. Antibodies (Ab) raised against the
XX polypeptides encoded by the genes are used for detection/identification
XX of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that
XX carry a gene-containing vector are used to select compounds that
XX modulate, regulate, induce or inhibit expression of the genes in plants,
XX animals or microorganisms other than P. luminescens and are able to alter
XX response or sensitivity to toxins and antibiotics produced by P.
XX luminescens. Cells transformed to express the genes are useful for
XX recombinant production of the proteins, particularly toxins and
XX antibacterials useful as insecticides, bactericides and fungicides. The
XX genes, proteins, vectors containing the genes and Ab are also useful
XX therapeutically (to treat microbial infection by bacteria or fungi that
XX are sensitive to P. luminescens-encoded toxins or antibiotics) and as
XX biopesticides. Other uses of the genes and the proteins are as virulence
XX factors and for identifying targets of human diseases for which P.
XX luminescens is a model (particularly plague and whooping cough). This
XX sequence represents one of the isolated P. luminescens genes
XX
XX Sequence 1008 BP; 377 A; 147 C; 191 G; 293 T; 0 U; 0 Other;
XX
XX Query Match 98.4%; Score 992; DB 10; Length 1008;
XX Best Local Similarity 99.0%; Pred. No. 6.2e-207; Indels 0; Gaps 0;
XX Matches 998; Conservative 0; Mismatches 10;
XX
XX 1 ATGGTTATACATTAATACACCTGATGATGAGTATTCACCCGTTGAAAGCAATA 60
DB 1 ATGGTTATACATTAATACACCTGATGATGAGTATTCACCCGTTGAAAGCAATA 60
QY 61 GCAGAGATATAGTACGTATCTAACTTTAGCAACAGATGAGGCTCATACGATCA 120
DB 61 GCAGAGATATAGTACGTATCTAACTTTAGCAACAGATGAGGCTCATACGATCA 120
QY 121 TATGGAATTGAAATCGAGCTAGAGAAATTAATATTAGCTTACGCTTGGCTGAAGTGT 180
DB 121 TATGGAATTGAAATCGAGCTAGAGAAATTAATATTAGCTTACGCTTGGCTGAAGTGT 180
QY 181 ATTCAATATGATCTAAACCTTCCTGATGACTATTATAGAAATTAAGAGACTGCTGAGGA 240

```

Db 181 ATTCAATAATGATCTAAACTTCTGATGACTATTATAGAAATPAAAGACCTGCTGAGACA 240
Qy 241 ATTATCAAGAAATATATGTTCTATATCTTTTCATCTGCACTATTAGGTGAAATGATGATCA 300
Db 241 ATTATCAAGAAATATATGTTCTATATCTTTTCATCTGCACTATTAGGTGAAATGATGATCA 300
Qy 301 ATTTCTAAAGATATGCGAAATGTTTATTAAGATGAATGATGATTTTGAAGTCAATAT 360
Db 301 ATTTCTAAAGATATGCGAAATGTTTATTAAGATGAATGATGATTTTGAAGTCAATAT 360
Qy 361 CCTGAAAACATTTGGAAATGCTCTGAGCTTGAATTAACATGAGTGCCTATTCTAGAT 420
Db 361 CCTGAAAACATTTGGAAATGCTCTGAGCTTGAATTAACATGAGTGCCTATTCTAGAT 420
Qy 421 GACGATTAATATATGACACTATATTTTCTCTGATACAGAAATTCACATGAGGAAAT 480
Db 421 GACGATTAATATATGACACTATATTTTCTCTGATACAGAAATTCACATGAGGAAAT 480
Qy 481 CACGATTAATATGACACTATATTTTCTCTGATACAGAAATTCACATGAGTGCCT 540
Db 481 CACGATTAATATGACACTATATTTTCTCTGATACAGAAATTCACATGAGTGCCT 540
Qy 541 GTTACTCTGCTGGAAGAGATTTTTCAAAAAACTTTTACAAATGATGAGGCTAAA 600
Db 541 GTTACTCTGCTGGAAGAGATTTTTCAAAAAACTTTTACAAATGATGAGGCTAAA 600
Qy 601 TCATTAGAGAAATTAATATGAGAGAAAAAACTTTCTTAAACCTTTCTTGACACCGCAG 660
Db 601 TCATTAGAGAAATTAATATGAGAGAAAAAACTTTCTTAAACCTTTCTTGACACCGCAG 660
Qy 661 AGATTACCTGATGAGAGATATGATTTTGGCTGACCAAGAAAGCGCTTAAATGAGAGA 720
Db 661 AGATTACCTGATGAGAGATATGATTTTGGCTGACCAAGAAAGCGCTTAAATGAGAGA 720
Qy 721 GTGAGTTTAAAGAACTTAAATATCAATCTAGAGATGATTTTCTATATGAGAGG 780
Db 721 GTGAGTTTAAAGAACTTAAATATCAATCTAGAGATGATTTTCTATATGAGAGG 780
Qy 781 GCTGCAAAAAGAAAGTATGTTCAATTATTAAGAGATCAAAAGGATACCTCCACAG 840
Db 781 GCTGCAAAAAGAAAGTATGTTCAATTATTAAGAGATCAAAAGGATACCTCCACAG 840
Qy 841 ACAGAGAGGAAAGTATGTTCAATTATTAAGAGATCAAAAGGATACCTCCACAG 900
Db 841 ACAGAGAGGAAAGTATGTTCAATTATTAAGAGATCAAAAGGATACCTCCACAG 900
Qy 901 TTATATAGTGTAGGCTAAGCCAAAAGACAGGTTACCTTACTCAAAATATATCTGAC 960
Db 901 TTATATAGTGTAGGCTAAGCCAAAAGACAGGTTACCTTACTCAAAATATATCTGAC 960
Qy 961 AATACAATGACGTTCTATGTTGGAATCTCATTAATAAATATATGA 1008
Db 961 AATACAATGACGTTCTATGTTGGAATCTCATTAATAAATATATGA 1008

RESULT 4
ACF67367_07/c
Continuation (8 of 57) of ACF67367 from base 700001 (Photobhabdus luminescens nucleotide
WP Sequence split into 57 fragments Locus ACF67367 Accession ACF67367
Fragment Name Begin End
WP ACF67367_00 1 110000
WP ACF67367_01 100001 210000
WP ACF67367_02 200001 310000
WP ACF67367_03 300001 410000
WP ACF67367_04 400001 510000
WP ACF67367_05 500001 610000
WP ACF67367_06 600001 710000
WP ACF67367_07 700001 810000
WP ACF67367_08 800001 910000
WP ACF67367_09 900001 1010000
WP ACF67367_10 1000001 1110000
WP ACF67367_11 1100001 1210000
WP ACF67367_12 1200001 1310000

WP ACF67367_13 1300001 1410000
WP ACF67367_14 1400001 1510000
WP ACF67367_15 1500001 1610000
WP ACF67367_16 1600001 1710000
WP ACF67367_17 1700001 1810000
WP ACF67367_18 1800001 1910000
WP ACF67367_19 1900001 2010000
WP ACF67367_20 2000001 2110000
WP ACF67367_21 2100001 2210000
WP ACF67367_22 2200001 2310000
WP ACF67367_23 2300001 2410000
WP ACF67367_24 2400001 2510000
WP ACF67367_25 2500001 2610000
WP ACF67367_26 2600001 2710000
WP ACF67367_27 2700001 2810000
WP ACF67367_28 2800001 2910000
WP ACF67367_29 2900001 3010000
WP ACF67367_30 3000001 3110000
WP ACF67367_31 3100001 3210000
WP ACF67367_32 3200001 3310000
WP ACF67367_33 3300001 3410000
WP ACF67367_34 3400001 3510000
WP ACF67367_35 3500001 3610000
WP ACF67367_36 3600001 3710000
WP ACF67367_37 3700001 3810000
WP ACF67367_38 3800001 3910000
WP ACF67367_39 3900001 4010000
WP ACF67367_40 4000001 4110000
WP ACF67367_41 4100001 4210000
WP ACF67367_42 4200001 4310000
WP ACF67367_43 4300001 4410000
WP ACF67367_44 4400001 4510000
WP ACF67367_45 4500001 4610000
WP ACF67367_46 4600001 4710000
WP ACF67367_47 4700001 4810000
WP ACF67367_48 4800001 4910000
WP ACF67367_49 4900001 5010000
WP ACF67367_50 5000001 5110000
WP ACF67367_51 5100001 5210000
WP ACF67367_52 5200001 5310000
WP ACF67367_53 5300001 5410000
WP ACF67367_54 5400001 5510000
WP ACF67367_55 5500001 5610000
WP ACF67367_56 5600001 564894

Query Match 98.4%; Score 992; DB 10; Length 110000;
Best Local Similarity 99.0%; Pred. No. 1.5e-206;
Matches 998; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 1 ATGGTTATCAATTAATTAACCTGATGATGAAAGTGATATCCACCGTTGAAAAGCAATTA 60
Db 6005 ATGGTTATCAATTAATTAACCTGATGATGAAAGTGATATCCACCGTTGAAAAGCAATTA 65946
Qy 61 GCAGAGATATATAGTACGATATCTAACTTTAAGCAAAAGAGATGAGGCTATACACATCA 120
Db 65945 GCAGAGATATATAGTACGATATCTAACTTTAAGCAAAAGAGATGAGGCTATACACATCA 65886
Qy 121 TATGAAATTTGAATATCGAGTAAAGAAATTAATATAGCTTAAGCTTTGGCTGTAGTGT 180
Db 65885 TATGAAATTTGAATATCGAGTAAAGAAATTAATATAGCTTAAGCTTTGGCTGTAGTGT 65826
Qy 181 ATTCATAATGATCTTAACTTCTGATGACTATTAATAAATAAAGAGACTGCTGAGAGA 240
Db 65825 ATTCATAATGATCTTAACTTCTGATGACTATTAATAAATAAAGAGACTGCTGAGAGA 65766
Qy 241 ATTTATCAAGAAATATATGTTCAATCTTTCATCTGCACTATTAGGTGAAATGATGATCA 300
Db 65765 ATTTATCAAGAAATATATGTTCAATCTTTCATCTGCACTATTAGGTGAAATGATGATCA 65706
Qy 301 ATTTCTAAAGATATGCGAAATGTTTATTAAGATGAATGATGATTTTGAAGTCAATAT 360
Db 65705 ATTTCTAAAGATATGCGAAATGTTTATTAAGATGAATGATGATTTTGAAGTCAATAT 65646

QY 361 CCTCAAAACATTTGGATGTTCTGTGAGCTGTGAATAAACATGAGCTTATTCAGAT 420
Db 65645 CCTCAAAACATTTGGATGTTCTGTGAGCTGTGAATAAACATGAGCTTATTCAGAT 65586
QY 421 GACGATTAATTTATGACCTATATTTTCTCTGTACAGAAATTCCTGGAGAAAT 480
Db 65585 GACGATTAATTTATGACCTATATTTTCTCTGTACAGAAATTCCTGGAGAAAT 65526
QY 481 CAACATCAATATGCGGAGAAATTTTAAATTAATGATTTCTTAACTTACCTTATCT 540
Db 65525 CAACATCAATATGCGGAGAAATTTTAAATTAATGATTTCTTAACTTATCTCT 65466
QY 541 GTAACCTCACTGGAGAGAGATTTTTCAAAACCTTTTCAATGATGATGAGCTTAA 600
Db 65465 GTAACCTCACTGGAGAGAGATTTTTCAAAACCTTTTCAATGATGATGAGCTTAA 65406
QY 601 TCATTTAGAAATTTATTTAGAGAAAAAACTTTCTTAACTTTCTTTCGACCCGAG 660
Db 65405 TCATTTAGAAATTTATTTAGAGAAAAAACTTTCTTAACTTTCTTTCGACCCGAG 65346
QY 661 AGATTAACCTGATGCGAGAAATGTTTGTGCTGACCAACAGAGCGCTTAAATGAGA 720
Db 65345 AGATTAACCTGATGCGAGAAATGTTTGTGCTGACCAACAGAGCGCTTAAATGAGA 65286
QY 721 GTGAGTTTAAAGAACTTAAATAACAATCTAGAAATGATTTTCTAATATGAGAGG 780
Db 65285 GTGAGTTTAAAGAACTTAAATAACAATCTAGAAATGATTTTCTAATATGAGAGG 65226
QY 781 GCTGCAAAAACAAAGATATAGTTTCAATTATTAAGAAGTACAAAGGGTAAAGCTTCACAG 840
Db 65225 GCTGCAAAAACAAAGATATAGTTTCAATTATTAAGAAGTACAAAGGGTAAAGCTTCACAG 65166
QY 841 ACAGAGCGGAAAGATATGTTGTTACGCCAGCTGCACTTACCTGAGAAATTTGCCGAATAT 900
Db 65165 ACAGAGCGGAAAGATATGTTGTTACGCCAGCTGCACTTACCTGAGAAATTTGCCGAATAT 65106
QY 901 TTATATAGTGTGAGGCTTAAGCCAAAAGACAGGGTAACTTTTCAAAATGATATCTGAC 960
Db 65105 TTATATAGTGTGAGGCTTAAGCCAAAAGACAGGGTAACTTTTCAAAATGATATCTGAC 65046
QY 961 AATACAATGACGGTTCATATGTTGGAACCTCATTTATTAATAATATATGA 1008
Db 65045 AATACAATGACGGTTCATATGTTGGAACCTCATTTATTAATAATATATGA 64998

RESULT 5
ACF65384_1/c
Continuation (2 of 6) of ACF65384 from base 100001. (Photorehabdus luminescens nucleotide
MP Sequence split into 6 fragments LOCUS ACF65384 Accession ACF65384
Fragment Name Begin End
MP ACF65384_0 1 110000
MP ACF65384_1 100001 210000
MP ACF65384_2 200001 310000
MP ACF65384_3 300001 410000
MP ACF65384_4 400001 510000
MP ACF65384_5 500001 530312

Query Match 98.4%; Score 992; DB 10; Length 110000;
Best Local Similarity 99.0%; Pred. No. 1.5e-206;
Matches 998; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 1 ATGGTTATTAACAATTAACCTGATGATAGTAGTATCCACCGGTTGAAAGCAATA 60
Db 66220 ATGGTTATTAACAATTAACCTGATGATAGTAGTATCCACCGGTTGAAAGCAATA 66161
QY 61 GCAGAGATATATGATGATATTAACCTTAAAGCAACAGATGAGGTCATACAGATCA 120
Db 66160 GCAGAGATATATGATGATATTAACCTTAAAGCAACAGATGAGGTCATACAGATCA 66101
QY 121 TATGAAATTTGAATATGAGCTTAAAGAAATATATATGCTTATGCTTGTATAGTGT 180
Db 66100 TATGAAATTTGAATATGAGCTTAAAGAAATATATATGCTTATGCTTGTATAGTGT 66041

QY 181 ATTCATATATGATCTAACTTCTGATGACTATTTAATAAGATTAAGAGACTGCTGAGAGA 240
Db 66040 ATTCATATATGATCTAACTTCTGATGACTATTTAATAAGATTAAGAGACTGCTGAGAGA 65981
QY 241 ATTTATCAAGATATATGCTTAACTTTTCACTGCACTATTAAGTGAATAATGCTGATCA 300
Db 65980 ATTTATCAAGATATATGCTTAACTTTTCACTGCACTATTAAGTGAATAATGCTGATCA 65921
QY 301 ATTTCTAAAGATATGCGGAAATGTTTTTAAAGATGAATGAACTTGAAGGTCATAT 360
Db 65920 ATTTCTAAAGATATGCGGAAATGTTTTTAAAGATGAATGAACTTGAAGGTCATAT 65861
QY 361 CCTCAAAACATTTGGAAATGTTCTGTGAGCTGTGAATAAACATGAGTCTTATTCAGAT 420
Db 65860 CCTCAAAACATTTGGAAATGTTCTGTGAGCTGTGAATAAACATGAGTCTTATTCAGAT 65801
QY 421 GACGATTAATTTATGACCTATATTTTCTCTGTACAGAAATTCCTGGAGAAAT 480
Db 65800 GACGATTAATTTATGACCTATATTTTCTCTGTACAGAAATTCCTGGAGAAAT 65741
QY 481 CAACATCAATATGCGGAGAAATTTTAAATTAATGATTTCTTATTAACCTTATCTGCT 540
Db 65740 CAACATCAATATGCGGAGAAATTTTAAATTAATGATTTCTTATTAACCTTATCTGCT 65681
QY 541 GTAACCTCACTGGAGAGAGATTTTTCAAAACCTTTTCAATGATGATGAGCTTAA 600
Db 65680 GTAACCTCACTGGAGAGAGATTTTTCAAAACCTTTTCAATGATGATGAGCTTAA 65621
QY 601 TCATTTAGAAATTTATTTAGAGAAAAAACTTTCTTAACTTTCTTTCGACCCGAG 660
Db 65620 TCATTTAGAAATTTATTTAGAGAAAAAACTTTCTTAACTTTCTTTCGACCCGAG 65561
QY 661 AGATTAACCTGATGCGAGAAATGTTTGTGCTGACCAACAGAGCGCTTAAATGAGA 720
Db 65560 AGATTAACCTGATGCGAGAAATGTTTGTGCTGACCAACAGAGCGCTTAAATGAGA 65501
QY 721 GTGAGTTTAAAGAACTTAAATAACAATCTAGAAATGATTTTCTAATATGAGAGG 780
Db 65500 GTGAGTTTAAAGAACTTAAATAACAATCTAGAAATGATTTTCTAATATGAGAGG 65441
QY 781 GCTGCAAAAACAAAGATATAGTTTCAATTATTAAGAAGTACAAAGGGTAAAGCTTCACAG 840
Db 65440 GCTGCAAAAACAAAGATATAGTTTCAATTATTAAGAAGTACAAAGGGTAAAGCTTCACAG 65381
QY 841 ACAGAGCGGAAAGATATGTTGTTACGCCAGCTGCACTTACCTGAGAAATTTGCCGAATAT 900
Db 65380 ACAGAGCGGAAAGATATGTTGTTACGCCAGCTGCACTTACCTGAGAAATTTGCCGAATAT 65321
QY 901 TTATATAGTGTGAGGCTTAAGCCAAAAGACAGGGTAACTTTTCAAAATGATATCTGAC 960
Db 65320 TTATATAGTGTGAGGCTTAAGCCAAAAGACAGGGTAACTTTTCAAAATGATATCTGAC 65261
QY 961 AATACAATGACGGTTCATATGTTGGAACCTCATTTATTAATAATATATGA 1008
Db 65260 AATACAATGACGGTTCATATGTTGGAACCTCATTTATTAATAATATATGA 65213

RESULT 6
AAK07349
ID AAK07349 standard; DNA; 1107 BP.
XX
XX AAK07349;
XX AC
XX 17-OCT-2003 (revised)
XX DT 21-MAY-1999 (first entry)
XX DE
XX Xenorhabdus nematophilus insecticidal toxin gene toxB4.
XX Toxin; toxB4; biological control; insecticide; ds.
XX Xenorhabdus nematophilus.
XX
XX W09903328-A1.
PN

```

XX 28-JAN-1999.
PD 17-JUL-1998; 98WO-AU000562.
XX PF 17-JUL-1997; 97AU-00008088.
XX PR 17-JUL-1997; 97AU-00008088.
XX (CSTR ) COMMONWEALTH SCI & IND RES ORG.
PA (CSTR ) COMMONWEALTH SCI & IND RES ORG.
XX East PD;
XX MPI; 1999-131733/11.
XX P-PSDB; AAM97810.
XX New insecticidal toxin genes - extracted from Xenorhabdus nematophilus
XX A24 and Photorhabdus luminescens.
XX Claim 1; Page 25-26; 48pp; English.
XX
XX This is the coding region of the insecticidally-active toxin gene, termed
XX toxb4, of Xenorhabdus nematophilus strain A24. It codes for a 368-amino
XX acid toxb4 protein (see AAM97810). The gene was isolated by functional
XX screening of a bacterial DNA library using Galleria mellonella fourth
XX instar larvae. The isolated toxb4 clone contained an insert of 1205 bp
XX (see AAX07351) that included the 1107 bp coding region. The invention
XX relates to the identification and isolation of polynucleotide molecules
XX encoding a new class of protein insecticidal toxins which are produced by
XX bacteria of the genera Xenorhabdus and Photorhabdus (see also AAX07350).
XX The polynucleotide molecules may be incorporated e.g. into insect-
XX specific viruses (e.g. entomopox and nuclear polyhedrosis viruses),
XX bacteria (e.g. Gracilicutes, Firmicutes, Temericutes and Mendosicutes),
XX protozoa, yeast and plants (particularly cereals such as wheat and
XX barley, vegetables such as tomato and potato, fruit trees such as citrus
XX and apple, tobacco and cotton) for control of pest insects. (Updated on
XX 17-OCT-2003 to standardise OS field)
XX
SQ Sequence 1107 BP; 424 A; 181 C; 199 G; 303 T; 0 U; 0 Other;
Query Match 65.9%; Score 663.8; DB 2; Length 1107;
Best Local Similarity 79.4%; Pred. No. 2.8e-135;
Matches 835; Conservative 0; Mismatches 162; Indels 54; Gaps 2;
QY 12 ATTAACACCTGATATAGAGAGTATCCACCCGTTGAAAGCAATATGCGAGATAT 71
DB 57 AGTAACGGCTGATGATTAAGGTAATATCAACCCGTTGAAAGCAATATGCGAGATAT 116
QY 72 AGTAGTATCTAATCTTTAAGCAAGATGAGGTCATACAGCATCATATGAAATTGA 131
DB 117 AATAGTGATCTAGATTCAGCAACAAATGAAGTCTATACGATTTGATGAAATTC 176
QY 132 ATATGAGCTTAAGAAATATATATAGCTTACGCTTGGCTGTATGATTCATATGT 191
DB 177 ATATGAGCTTAAGAAATATATATAGCTTACGCTTGGCTGTATGATTCATATGT 236
QY 192 ATCTAATCTTCCGATGATATATATAGAAATGAAGACTGCTGAGAAATTTATCAAGA 251
DB 237 CTCCTAACCTCCAGAGACTATATATGAATGAATGAATGAATGAATTTATCAAGA 296
QY 252 ATATATGCTAATCTTTGATCTGCACTATATAGGTAAGTAAGTGTATCAAAATTTCAAAGA 311
DB 297 ATACATGCTAATCTTTTATCTGCACTATATGAGTATGAGTATGATCAAAATTTCAAAGA 356
QY 312 TATGCAAAATGCTTTTATATAGAAATGAATGATTTTGAAGTCAATATCTCAAAACAT 371
DB 357 TATGCAAAATGATTTTATCCAGAAAGAACTGAGTTTGAAGTCAACGCTTAAAAATAC 416
QY 372 TTGGAATGTTCTGAGCTTGAAGAAATGAATGAGTGTATTTCAAGTACGATTAAT 431
DB 417 CTGGATATCTCTGATCTTGAGAAATGAATGAATGAATGAATGAATGAATGAAT 476
QY 432 ATTACAGATATATTTTCTCTGTACAGGAATTCACATGAGGAAATCAATCAAA 491
DB 477 ATTACAGATATATTTCTTCTGCTTCAAGAACTTCAATGAGGCAATCAATCAAA 536

```

```

QY 492 TGCCGCAAGATTTTAAATTAATGATTTCTTATTAACCTTATGCTGTACTACT 551
DB 537 TGCAGCAAAATTTTAAAGTAATGATTTTCTTATCTTATCTGCTGAACATCACT 596
QY 552 GGGAAAGAGATTTTTCACAAAACCTTTACAAAGATTTAGAGCTAAATCATTTAGAGA 611
DB 597 GGGAAAAAGATTTTTCACAAAACCTTTACAAAGATTTAGAGCTAAATCATTTAGAGA 656
QY 612 TTATATGAGAAAAAAACCTTTCTAAACCTTTCTTTCGACACCGCAGAGATTACTGA 671
DB 657 TTATATGAGAAAAAAACCTTTCTAAACCTTTCTTTCGACACCGCAGAGATTACTGA 716
QY 672 TGGCAAAATGAGTATTTGGCTGACCAACAGAGCGCTTAA----- 714
DB 717 TGGCAGAAACAGGCTACTTGGCCGGTCCAAACAAAGCGCTTAAATTGCCAACACGCTTTC 776
QY 715 -----TGGAGGTGAGTTTAAAGACTTAA 740
DB 777 TACAGCAAAACGCTACAGACGCTTCACTTAATTTGAGAGTTAGTTGCAAAAACCTTAG 836
QY 741 AATTAACAAATCTAGAAATGATTTTCTATATGAGAAAGGCTGCAAAAACAAAGTATAG 800
DB 837 AGATTAACCAATCCAGAAATCATTTATGAAATGAGATGATGCTGCAAAAACGAAATATAG 896
QY 801 TTCAATTTTAAAGAGTTACAAAAGGTTACGCTCCACAGACAGCAGCGAAAGTATTGG 860
DB 897 TTCAATTTTAAAGAGTTACAAAAGGTTACGCTCCACAGACAGCAGCGAAAGTATTGG 956
QY 861 TACAGCGAGTGGAGTAACTGGAAGAAATTTGCCGAAT-ATTATATAGTGGAGCT 917
DB 957 TACAAAAGCGCGATTAATCTTGAAAACCTGCAAGGTGAGATTTATATATAGTAACT 1016
QY 918 AAGCAAAAAGACAGGGTAACTTTACTCAAAATGATCTGACATACATAGACGGTTCA 977
DB 1017 AAGCAAAAAGACAGGGTAACTTTCTCATTAATATCTGACCAAAATATATGAGATCA 1076
QY 978 TAGTGTGAACTCATTTATTAATAATATATGA 1008
DB 1077 AAGTGTGAACTCATTTACCAAAAATATATGA 1107
RESULT 7
AAX07351
ID AAX07351 standard; DNA; 1205 BP.
XX
XX AAX07351;
AC 17-OCT-2003 (revised)
DT 21-MAY-1999 (first entry)
XX
XX Xenorhabdus nematophilus insecticidal toxin gene toxb4.
XX
XX Toxin; toxb4; biological control; insecticide; ds.
XX
XX Xenorhabdus nematophila.
OS
XX
XX Key Location/Qualifiers
FH CDS 17..1123
FT /+tag= a
FT /note= "the coding region is specifically claimed in
FT Claim 1"
XX
XX WO9903328-A1.
XX
XX 28-JAN-1999.
XX
XX 17-JUL-1998; 98WO-AU000562.
XX
XX 17-JUL-1997; 97AU-00008088.
XX
XX (CSTR ) COMMONWEALTH SCI & IND RES ORG.
XX

```

OY	12	TTTAAACCGTATGATGAGAGGATATCACCGCTGAAAAAGCAATGCGAGGAT	71
Db	73	AGTAAGCGCTGATGATTAAGTGAATATCACCCGTTGAAAACAAATGCGGAGAT	132
OY	72	AGTAGTATACTMAACTTTAAGCAACAGATGAGGGTCATACGACATCATGGAATTGA	131
Db	133	AATAGGTGACTAGAAATTCAGACCAACAAATBAAAGTCATACAGGATTTGATGAATTGC	192
OY	132	ATATGAGCTAAAGAAATATATTTAGCTTAACGCTTGGCTGTAAGTGGTATTCATATGT	191
Db	193	ATATGAGCTAAAGAAAGTATATATGCAATATGCTTTAGCGGTAGGTGATTCATATGT	252
OY	192	ATCTAAACTTCCTGATGACTATTTAAGAAATTAAGAGATGCTGAGAGAAATTTACAAGA	251
Db	253	CTCTCAACTTCAGAAAGACTATTTAATAAATTAAGGATACACAGGTGAATTTATCABA	312
OY	252	ATATATGCTATCTTCTTCATCTGCACCTATTAGGTGAATAAGTGATCAATTTCTTAAGA	311
Db	313	ATACATGCTATCTTTTATCTGCACATATGGGTGAATGGTGAATCAATTTCTTAAGA	372
OY	312	TATGCGAAATGGTTTTATTAAGATGAACGTGAATTTGAAGTCAATATCTCAAAACAT	371
Db	373	TATGCGAAATGATTTTATCCAGAACGAATGAGGTTTGAGGTCACATCTTAAAAATAC	432
OY	372	TTGGAATGTTCTGAGCTTGAAAAATPAACCATTTAGTGCTTATTCAGATGAGATPAAT	431
Db	433	CTGGGATATTTCTGATCTTGAAGATTAACATATTGGAAGATTTATCGAATGAAGATPAAT	492
OY	432	ATTAGCATATATTTTTTCTGTACAGAGAAATTCACGTGAGAGAAATCAACAATCAAA	491
Db	493	ATTAGCATATATTTCTTGTCTTCACAAAGACTCCAAATGAGGCAAAATCAACAATCAAA	552
OY	492	TGCCGCAATTTTTTAAATTAATGATTTCTTATTTACCTTATCTGCTGTAACCTCACT	551
Db	553	TCCAGCAATTTTTTAAAGTATGATTTTACTATCTTATCTGCTGTAACATCACT	612
OY	552	GGGAGGAGAGATTTTTCAAAAAACCTTTACATGATTAAGGCGTAATTCATTAAGAA	611
Db	613	GGGAAAAAGATTTTTTCAAAAAATTTTACATAGGTCTAGAACTTAATCTATTAGAA	672

DR P-PSDB; AAR68961.

RESULT 8	
AAQ80839	
ID	AAQ80839 standard; DNA; 1272 BP.
XX	
XX	
AC	AAQ80839;
XX	
XX	
DT	12-FEB-2004 (revised)
DT	27-AUG-2003 (revised)
DT	09-AUG-1995 (first entry)
XX	
XX	
DE	Xenorhabdus nematophilus strain A24 insecticidal toxin DNA.
XX	
KM	Bacterium; insecticide; pesticide; toxin; ss.
XX	
OS	Xenorhabdus nematophila.
XX	
XX	
FH	Key
FT	-35_signal
FT	Location/Qualifiers
FT	44..49
FT	/*tag= a
FT	66..71
FT	/*tag= b
FT	71..79
FT	/*tag= c
FT	83..919
FT	/*tag= d
FT	
PN	WO9500647-A1.
XX	
XX	
PD	05-JAN-1995.
XX	
PF	24-JUN-1994;
XX	94WO-AU000348.
PR	25-JUN-1993;
XX	93AU-00009638.
PA	(CSIR) COMMONWEALTH SCT & IND RES ORG.
XX	
PI	Smigielski AJ, Akhurst RJ;
XX	
DR	WPI: 1995-052084/07.
DR	P-PSDB; AAR68961.

Qy	441	ATATTTTTCCTGTACAGGAATTCACCTGAGAGAAATCAACATCAAAATGCCGAG	500
Db	1068	TTTTTTTATTACCAATTAATTAATTCAAAATTTAAAAAATTAATAAAAAATTTTAAAT	1000
Qy	501	ATTTTAAATTAATGATTCCTATTTTACCTTATCTGCTGACTTCACTGGAGAGAG	560
Db	1008	AATTAATTTTAAAAAATTAATATATATTTTATCATTAAATAAAAAATTAATAAT	949
Qy	561	GATTTTTCAAAAAATTTTACATGATTAAGGCTAAATCATAGAGAAATTAATGA	620
Db	948	TATTTTAAAAAATTTTATTCAAAAAACAAAAAAATTAATTAATTAATTAATAA	889
Qy	621	GAGAAATAAATCTTCTAA	639
Db	888	ATTACAAAATTTTATACA	870

RESULT	11
ID	AA60472
XX	AA60472 standard; DNA; 4590 BP.
AC	AA60472;
XX	
DT	25-MAR-2003 (revised)
DT	24-AUG-1991 (first entry)
XX	
DE	Sequence encoding the ring-infected Erythrocyte Surface Antigen (RESA)
XX	
KW	Malaria vaccine; antigen; epitope; ss.
XX	
OS	Plasmodium falciparum.

PH	Key	Location/Qualifiers
FT	exon	801..995
FT		/*tag= a
FT	exon	1199..4225
FT		/*tag= b
XX		
PN	W08601802-A.	
XX		
PD	27-MAR-1986.	
XX		
PF	11-SEP-1984;	84AU-00007066.
XX		
PR	11-SEP-1984;	84AU-00007066.
PR	11-SEP-1984;	84AU-00007067.
PR	10-SEP-1985;	85AU-00047326.
XX		
PA	(HALL-) HALL INST MEDICAL RES WALTER & ELIZA.	

PI Kemp DJ, Anders R, Coppel RL, Brown G, Saint RB, Comman AF,
XX
XX WPI; 1986-094065/14.
DR
DR P-PSDB; AAP60569.
XX
PT DNA coding for Plasmodium falciparum antigens - expressing
PT polyepitope(s) having antigenicity of RBSA or FRA antigens of P
PT falciparum.
PT

PS Claim 4; Fig 1; 55pp; English.

CC The inventors claim a novel DNA molecule which comprises a nucleotide
CC sequence corresp. to all or a portion of the base sequence coding for
CC (AAN60472) or F1RA (AAN60473). RISA and F1RA have antigenicity suitable
CC for providing protective immunity against *Plasmodium falciparum* malaria
CC infections. (Updated on 25-MAR-2003 to correct PA field.)

SQ Sequence 4590 BP; 1933 A; 437 C; 673 G; 1547 T; 0 U; 0 Other;

Query Match	5.0%	Score 50.8	DB 1	Length 4550
Best Local Similarity	43.6%	Pred. No. 0.24		
Matches 226; Conservative	0	Mismatches 292	Indels 0	Gaps 0

QY	1	ATGGCTATACAACTTAACACCTGATGATGAAAGTGGATATATCCACCCGTTGAATAACCAATA	60
Db	3587	ATTGAGGAAATATATTATGACATGGAATTAAGCAATTAATACAGAAATGAAAACCAAAAT	3646
QY	61	GCAGAGATATATGATGATGATATCTTAAACCTTTAAGCAAAACAGATGAGGCTCATACAGATCA	120
Db	3647	GAAATATGACAGAAACATGTACCACTATATGCTGAAGAAAATGTAGAAACATGATGCTGAA	3706
QY	121	TATGGAATTTGAATATCGAGCTTAAGAAAATAATATTAGCTTACGGTTTGGCTGTAAAGGT	180
Db	3707	GAAAAATGTAGAACATGATGCTGAAAGAAAATGTAAACATGATGCTGAAAGAAAATGTAGAA	3766
QY	181	ATTCTAATATGATCTTAAACCTTCCGATGACATTTATTAAGATTAAGAGACTGCTGACAGA	240
Db	3767	CATGATGCTGAAGAAAATGTAGAACATGATGCTGAAAGAAAATGTAGAAAGAAAATGTTGAA	3826
QY	241	ATTATTAACAAGAAATATATGCTTAAATCTTTCATCTGCACTATTAGGTGAAAAATGTGATCA	300
Db	3827	GAAGTAGAAGAAAATGTAGAAAGAAAATGTAAAGAAAATGTAGAAAGAAAATGTTGAAGAA	3886
QY	301	ATTTCTAAGATATGCGAAATGTTTTTATTAAGATGAACCTGATTTTGAAGTCAATAT	360
Db	3887	GTAGAGAAAATGTAGAAAGAAAATGTAGAAAAAATGTAGAAAGAAAATGTTGAAGAAAAT	3946
QY	361	CCTCAAAACATTTGGAATGTTCCCTGAGCTGAAAAATTAACCATTTGAGTGTATTATTCAGAT	420
Db	3947	GTTGAGAAAATGTTGAAGAAAATGTAGAAAAAATGTAGAAAGAAAATGTTGAAGAAATAT	4006
QY	421	GACGATTAATATTATGACATTAATTTTTTCTCTGACAGAAATTTCACTTGAAGAAAAT	480
Db	4007	GATGAGAAAATGTTGAAGAAAGTGAAGAAAATGTAGAGAAAATGTAGAAAGAAAATGTT	4066
QY	481	CAACATCAAAATGCCGCAAGTTTTTAAATTAAATTGA 518	
Db	4067	GAAGAAAATGTAGAAAGAAAATGTTGAAGAAAGTGAAGAA 4104	

RESULT 12
ADL08108
ID ADL08108 standard; DNA; 186971 BP.

AC ADL08108;

DT 20-MAY-2004 (first entry)

DE Human gene associated with low HDL-C APOA1.

KN Human; ds; SNP; single nucleotide polymorphism;

KW coronary artery disease; gene.

Os Homo sapiens.

FH	Key	Location/Qualifiers
1	1	1
2	2	2
3	3	3
4	4	4
5	5	5
6	6	6
7	7	7
8	8	8
9	9	9
10	10	10
11	11	11
12	12	12
13	13	13
14	14	14
15	15	15
16	16	16
17	17	17
18	18	18
19	19	19
20	20	20
21	21	21
22	22	22
23	23	23
24	24	24
25	25	25
26	26	26
27	27	27
28	28	28
29	29	29
30	30	30
31	31	31
32	32	32
33	33	33
34	34	34
35	35	35
36	36	36
37	37	37
38	38	38
39	39	39
40	40	40
41	41	41
42	42	42
43	43	43
44	44	44
45	45	45
46	46	46
47	47	47
48	48	48
49	49	49
50	50	50
51	51	51
52	52	52
53	53	53
54	54	54
55	55	55
56	56	56
57	57	57
58	58	58
59	59	59
60	60	60
61	61	61
62	62	62
63	63	63
64	64	64
65	65	65
66	66	66
67	67	67
68	68	68
69	69	69
70	70	70
71	71	71
72	72	72
73	73	73
74	74	74
75	75	75
76	76	76
77	77	77
78	78	78
79	79	79
80	80	80
81	81	81
82	82	82
83	83	83
84	84	84
85	85	85
86	86	86
87	87	87
88	88	88
89	89	89
90	90	90
91	91	91
92	92	92
93	93	93
94	94	94
95	95	95
96	96	96
97	97	97
98	98	98
99	99	99
100	100	100

```

...
/*tag= a
...

```

[illegible][illegible]

PD 04-MAR-2004.

04-SEP-2002; 2002US-00235192.

PR 04-SEP-2002; 2002US-00235192.

PA (VITI-) VITIVITY INC.

AA
PI
Mccarthy J:

WPI: 2004-214170/20.

AA Determining whether a subject has, or is at risk of developing, an

Sequence 3738 BP; 1705 A; 258 C; 542 G; 1233 T; 0 U; 0 Other;

Query Match	Score	DB 3	Length
5.0%	50.4	DB 3	3738

Matches 271; Conservative 0; Mismatches 331; Indels 2; Gaps 2;

QY	170	CTGTAAGGCGATTTCATAATGATATTAACCTCCGATGACATATATATAGAAATAAGGA	229
Db	2537	CTACTAATAGTAATGATATGTAATGATAGTAATATACATATATATATATATATAT	2596
QY	230	CTGCTGAGAAATTTATCAAGAAATATATGTCCTAATCTTTCATCTGACACTATTAGTGAAA	289
Db	2597	ATAATATATATATATATATATATATATATATATATATATCTGTCATATTTTCAACAAATGAAAAA	2656
QY	290	ATGCTGATCCAAATTTCTAAAGTATANGCAATGCTTTTATATAGAAATGACCTGAGTTTGG	349
Db	2657	TTTACGATATGCTAAATATGGATATATATATATATATATATATATATATATATATATAT	2716
QY	350	AAGCTCAATATCTCTCAAAACATTTGGATGTTCTGAGCTTGAAATATTAACCATTGAGTG	409
Db	2717	GTCATTCGAT	2775
QY	410	CTTATTCAGATGACGAT	469
Db	2776	AATATATGACATATTTGAT	2835
QY	470	TGGAGGAAATATCAACATCAATATGCCGAAAGTTTTTAAATATATATATATATATATAT	528
Db	2836	AAT	2895
QY	529	ACCTTATCTGCTGTAACCTCACTGGGAGAGAGATTTTTTCAAAAACTTTTACAAATGGA	588
Db	2896	TCTAT	2955
QY	589	TTAAGGCTAATATCAATTAGAGATATATATATATATATATATATATATATATATATATAT	648
Db	2956	TTGAGAGAAAT	3015
QY	649	CGACCAACGCGAGATTAACCTGATGGCAAGATATATATATATATATATATATATATATATAT	708
Db	3016	CACAAAGGAAT	3073
QY	709	CCTAATATGAGATGATTTTAAAGACCTTAAATAATCAATCTAGAGATGATTTTCT	768
Db	3076	AAATGTTTTTTTTTAAATTTTCCAAATATATATATATATATATATATATATATATATATGA	3133
QY	769	AATA 772	
Db	3136	AAAA 3139	

RESULT 14
ABK28221
ID ABK28221 standard; DNA; 11394 BP,
vv

DT 23-APR-2002 (first entry)

DE DNA transcription associated genomic DNA #48.

KM DNA transcription associated gene; peptide nucleic acid; pNA-oligomer;
 KM PNA; cytosine methylation state; SNP; retroviral infection; gene; ds;
 KM single nucleotide polymorphism; adenosine deaminase deficiency; cancer;
 KM viral infection; Searcy syndrome; haematological disorder; tuberculosis;
 KM immunological disorder; Werner syndrome; developmental disorder;
 KM psoriasis; Rieger's syndrome; neurological disorder; erythropoietic;
 KM neurodegenerative disorder; Waardenburg syndrome; Niemann-Pick disease;
 KM myeloid/plastic syndrome; myocardial infarction; hypertension; arthritis;
 KM angiogenesis; congenital heart disease; HDR syndrome; gene therapy;
 KM polyglutamine disorder; solid tumour.

XS unidentified

XX WO200192565-A2.
 PN
 XX
 PD 06-DEC-2001.
 XX
 PF 06-APR-2001, 2001WO-EP003973.
 PF

PR 06-APR-2000; 2000DE-01019058.
PR 07-APR-2000; 2000DE-01019173.
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX

PA (EPIG-) EPIGENOMICS AG.
XX
XX
PI Olek A, Piepenbrock C, Berlin K,
XX
DR WPI; 2002-090046/12.
XX

PT New nucleic acids or oligomers, useful for diagnosing or treating diseases associated with DNA transcription, e.g. immunological disorders, Werner syndrome, psoriasis, myocardial infarction, solid tumors or cancer.

The invention relates to a nucleic acid, which comprises a segment of the chemically pretreated DNA of genes associated with DNA transcription from one of 346 sequences, and an oligomer, in particular an oligonucleotide or peptide nucleic acid (PNA)-oligomer that hybridises to or is identical to the chemically pretreated DNA of genes associated with DNA transcription. The set of oligomer probes are useful for detecting the cytosine methylation state and/or single nucleotide polymorphisms (SNPs) in a chemically pretreated genomic DNA. The nucleic acids are useful for diagnosing or treating diseases associated with DNA transcription (particularly with the methylation status), e.g. adenosine deaminase deficiency, viral infection, retroviral infection, Searcy syndrome, haematological disorders, immunological disorders, Werner syndrome, tuberculois, developmental disorders, psoriasis, Rieger's syndrome, neurological disorders, neurodegenerative disorders, Waardenburg syndrome, Niemann-Pick disease, myelodysplastic syndrome, myocardial infarction, hypertension, angiodogenesis, erythropoiesis, congenital heart disease, HDR syndrome, arthritis, polyglutamine disorders, solid tumours or cancer. Sequences ABK28127-ABK28472 represent DNA transcription associated genomic DNA molecules of the invention. Note: The sequence data for this patent did not form part of the printed specification but was obtained in electronic format directly from the European Patent Office

SO	Sequence	11394	BF,	3189	A;	200	C,	2921	G;	5084	T;	0	U;	0	Other;
	Query Match		4.9%;	Score	49.8;	DB	6;	Length	11394;						
	Best Local Similarity		50.6%;	Pred. No.	0.46;										
	Matches	120;	Conservative	0;	Mismatches	117;	Indels	0;	Gaps	0					

OY	124	GAATTTGAATTCAGCTCAGAAATATATTAGCTCAGCTTGAGCTGATGTAAT	183
Db	318	GGAATTCGATTGAGTGGTGAAGATATTAGAAAGAAATTTATGTTAGTTTATAGATATATAT	377
OY	184	CATATATGATCTAACTTCCTGATGACATATTATAAAGATTAAGAGACTGCTGAGCAAT	243
Db	378	ATATATATATATATATATAAATGTGTATATATGTATATAAATATTTATTAATAATGTGTATAT	437
OY	244	TATCAGAAATATATGCTCTAATCTTTCATCTGCATATTAGGTGAAATGCTGATCAAT	303
Db	438	AAGGTTTGATTTTATTATATATATTTATTTGTTTGTAGTTGAGAGGTTTGAATAATA	497
OY	304	TCTAAGATATAGCGAAATGTTTTTATATAGAAATGAACAGATTTTGAAGCTCAAT	360
Db	498	TGATATTTTATTAGTATGATATGATTTATTAGTGCTTTAGATGTTTTTGGTTTTTAATAT	554

RESULT 15
AAA70105

ID AAA70105 standard; DNA; 5940 BP.
 AC AAA70105;
 DT 07-NOV-2000 (first entry)
 XX Plasmodium falciparum chromosome 2 related DNA sequence SEQ ID NO:238.
 DE Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
 KW antimalarial; malaria; protozoacide; infection; insecticide; ds.
 OS Plasmodium falciparum.
 MO200025728-A2.
 XX
 XX
 XX 11-MAY-2000.
 XX
 XX 05-NOV-1999; 99WO-US026796.
 XX
 XX 05-NOV-1998; 98US-0107131P.
 PR
 PA (HOPE/) HOFFMAN S.
 PA (CARU/) CARUCCI D.
 PA (GARD/) GARDNER M.
 PA (VENT/) VENTER J C.
 XX
 PI Hoffman S, Carucci D, Gardner M, Venter JC;
 XX
 DR WPI; 2000-365347/31.
 PT Proteins encoded by chromosome 2 of the human malarial parasite,
 PT Plasmodium falciparum, useful as antimalarial vaccines and in the
 PT diagnosis of P. falciparum infection.
 XX
 PS Disclosure: Page 460-462; 577pp; English.
 XX
 XX The present invention describes proteins and their fragments (I) encoded
 CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
 CC Also described are: (i) nucleotide sequences (ii) encoding (I); and (2)
 CC vaccines against P. falciparum infection comprising (i) or (ii). (i) and
 CC (ii) are useful for the development of vaccines against P. falciparum
 CC infection. (i) and polyclonal antisera or a monoclonal antibody raised to
 CC immunogens comprising the sequences of (i), are useful in the detection
 CC of infection with P. falciparum. Furthermore, (i) (especially when they
 CC are refined or secreted or membrane proteins) can aid the identification
 CC of drugs to treat or prevent P. falciparum infection, or they can be used
 CC to identify drug resistance in P. falciparum. Sequencing of the
 CC Plasmodium chromosome 2 and the subsequent identification of proteins
 CC encoded by it will help to expand our understanding of parasite biology,
 CC a process hampered by the complexity of the parasitic lifecycle, and
 CC provide new targets for vaccine and drug development. Parasite resistance
 CC to drugs and mosquito resistance to insecticides have led to a resurgence
 CC of malaria in many parts of the world, and there is a pressing need for
 CC vaccines and new drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352
 CC represent nucleotide and protein sequences given in the present
 CC invention, but which are not specifically mentioned within the
 CC specification
 CC
 SQ Sequence 5940 BP; 3106 A; 343 C; 879 G; 1612 T; 0 U; 0 Other;

Query Match 4.9%; Score 49.2; DB 3; Length 5940;
 Best Local Similarity 43.3%; Pred. No. 0.56;
 Matches 284; Conservative 0; Mismatches 368; Indels 4; Gaps 1;

QY 175 AGTGTATTCATATGATGATCTTAACCTCTGATGACTATTTAAGATTAAGAGACTGCT 234
 DB 229 AATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 288
 QY 235 GAGGAGATTTCAAGATATATGCTATCTTTCATCTGACATTAAGTGAAGTGT 294
 DB 289 AATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 348
 QY 295 GATCAATTTCTTAAGATATGCAATGCTTTTATTAAGATGAACCTGATTTGAAGGT 354

DB 349 GATGAAATATCAAAAAATATATATACATTAAGACATGAGTGAAGAAACACCTTAAGAT 408
 QY 355 CAATATCCTCAAAACATTTGGAATGTTCCAGGCTGAAAATTAACGATGAGTCTAT 414
 DB 409 ACATTAAAGTCATTAGTTCGTTTCGATTAATAATGGAATTAAGAAATTAAGTGA 468
 QY 415 TCAGATGACGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 474
 DB 469 GAATTTAGAAAAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTA 528
 QY 475 GAAATCAACATCAATCAATGCGCAGATTTTAAATTAATTAATTAATTAATTAATTAATTA 534
 DB 529 AATTAATTAAGAAAAAG-----AAGATTTGTTAAACAAAAAATGATATGCTTAATGA 584
 QY 535 TCTGCTGAATCTTCACTGCGAAGAGATTTTCAAAAACTTTCAATGATTAAG 594
 DB 585 AAAAAAATCTTTTACAGAAAAAAGATTAATTAATTAATTAATTAATTAATTAATTAATTA 644
 QY 595 GCTAATCATTAAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 654
 DB 645 TGAAGAAAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 704
 QY 655 CCGAGAGATTACCTGATGCGAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 714
 DB 705 TTTAGAAAAAATTAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 764
 QY 715 TGGAGAGGATTTTAAAGAACTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 774
 DB 765 TCTAGAAAAAATTAAGAAATTAAGAAATTAAGAAATTAAGAAATTAAGAAATTAAGAAATTA 824
 QY 775 GAAAGGCTGCAAAACAAAGTATAGTTCATTTAATAAGATTAAGATTAAGATTAAGATTAAG 830
 DB 825 AAATTTTATCAAAAGAAATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTA 880

Search completed: November 21, 2004, 06:27:15
 Job time : 546 secs

THIS PAGE BLANK (USPTO)

THIS PAGE

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_nzp model

Run on: November 21, 2004, 08:42:25 ; Search time 29.5 Seconds
(without alignments)
4532.105 Million cell updates/sec

Title: US-10-617-962-2

Perfect score: 1735

Sequence: 1 atggttatcaattacacc.....ctcattataaataatata 1008

Scoring table:

BLOSUM62	
Xgapop 10.0 , Xgapext 0.5	
Ygapop 10.0 , Ygapext 0.5	
Fgapop 6.0 , Fgapext 7.0	
Delop 6.0 , Delext 7.0	

Searched: 478139 segs, 66318000 residues

Total number of hits satisfying chosen parameters: 956278

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODE=frame_nzp.model -DEV=xlp
-O=/cgn2_1/USPTO.spool.p/US10617962/rnat.17112004.163923.9771/app.query.faeta.1.1159
-DB=Issued Patents AA -QFMT=fastan -SUFFIX=rai -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=BITS -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10617962.0CGN1.1.33.0rnat.17112004.163923.9771 -NCPU=6 -ICPU=3
-NO MMAP -LARGEOUTERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued Patents AA:*
1: /cgn2_6/prodata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/prodata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/prodata/1/1aa/6A.COMB.pep:*
4: /cgn2_6/prodata/1/1aa/6B.COMB.pep:*
5: /cgn2_6/prodata/1/1aa/PCTUS.COMB.pep:*
6: /cgn2_6/prodata/1/1aa/backfill1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1734	99.9	335	4	US-09-463-048A-4
2	1276	73.5	368	4	US-09-463-048A-3
3	1058	61.0	383	4	US-08-569-168-7
4	940.5	54.2	278	2	US-08-569-168-2
5	105.5	6.1	816	4	US-09-328-352-6677
6	101.5	5.9	600	4	US-09-248-795A-23971
7	96.5	5.6	293	1	US-08-628-291-4
8	96.5	5.6	293	2	US-09-128-722-4
9	96.5	5.6	317	1	US-08-628-291-12
10	96.5	5.6	317	2	US-09-128-722-12
11	96	5.5	623	4	US-09-538-092-119
12	96	5.5	912	2	US-08-951-871-2

13	96	5.5	3168	4	US-09-489-039A-14067	Sequence 14067, A
14	95.5	5.5	680	4	US-09-583-110-3191	Sequence 3191, Ap
15	95.5	5.5	915	1	US-08-328-322-5	Sequence 5, Appl1
16	94.5	5.4	400	4	US-09-248-796A-16641	Sequence 16641, A
17	94.5	5.4	696	3	US-08-899-437-23	Sequence 23, Appl
18	94.5	5.4	696	3	US-09-126-121-23	Sequence 23, Appl
19	94.5	5.4	753	4	US-09-543-681A-5022	Sequence 5022, Ap
20	94.5	5.4	863	2	US-08-666-271-2	Sequence 271, Ap
21	94	5.4	325	4	US-09-134-000C-4195	Sequence 4195, Ap
22	94	5.4	534	4	US-09-538-092-673	Sequence 673, App
23	94	5.4	910	4	US-09-623-326-7	Sequence 7, Appl1
24	94	5.4	1388	2	US-08-685-576-1	Sequence 1, Appl1
25	92.5	5.3	688	4	US-09-248-796A-15231	Sequence 15231, A
26	92.5	5.3	1572	2	US-08-290-731C-5	Sequence 5, Appl1
27	92.5	5.3	1596	3	US-09-356-952-3	Sequence 3, Appl1
28	92	5.3	598	2	US-08-937-540-2	Sequence 2, Appl1
29	92	5.3	598	4	US-09-398-395A-26	Sequence 26, Appl
30	92	5.3	598	4	US-09-887-586A-26	Sequence 26, Appl
31	92	5.3	598	4	US-09-895-752-26	Sequence 26, Appl
32	92	5.3	598	4	US-09-903-012B-26	Sequence 26, Appl
33	92	5.3	598	4	US-09-900-797-26	Sequence 26, Appl
34	91.5	5.3	399	4	US-09-107-532A-4714	Sequence 4714, Ap
35	91.5	5.3	475	4	US-09-248-796A-17531	Sequence 17531, A
36	91.5	5.3	485	4	US-09-639-245-8	Sequence 8, Appl1
37	91	5.2	311	4	US-09-248-796A-16692	Sequence 16692, A
38	91	5.2	310	4	US-09-623-326-8	Sequence 8, Appl1
39	90.5	5.2	933	4	US-09-543-681A-6487	Sequence 6487, Ap
40	90	5.2	415	4	US-09-543-681A-6746	Sequence 6746, Ap
41	90	5.2	447	3	US-08-961-083-182	Sequence 182, App
42	90	5.2	447	4	US-09-536-784-182	Sequence 182, App
43	90	5.2	484	4	US-09-468-656A-6	Sequence 6, Appl1
44	90	5.2	1039	4	US-09-583-110-5226	Sequence 5226, Ap
45	90	5.2	1388	2	US-08-685-576-4	Sequence 4, Appl1

ALIGNMENTS

RESULT 1
US-09-463-048A-4
; Sequence 4, Application US/09463048A
; Patent No. 6630619
; GENERAL INFORMATION:
; APPLICANT: Commonwealth Scientific and Industrial Research Organisation
; APPLICANT: EAST, Peter David
; TITLE OF INVENTION: Toxin Genes from the Bacteria Xenorhabdus nematophilus and Photob
; FILE REFERENCE: 050179-0076
; CURRENT APPLICATION NUMBER: US/09/463,048A
; CURRENT FILING DATE: 2002-12-13
; PRIOR APPLICATION NUMBER: PCT/AU98/00562
; PRIOR FILING DATE: 1998-07-17
; PRIOR APPLICATION NUMBER: PO 8088
; PRIOR FILING DATE: 1997-07-17
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 335
; TYPE: PRT
; ORGANISM: Photobdus luminescens
US-09-463-048A-4

Alignment Scores:

Pred. No.:	2.5e-174	Length:	335
Score:	1734.00	Matches:	335
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	99.94%	Indels:	0
DB:	4	Gaps:	0

US-10-617-962-2 (1-1008) x US-09-463-048A-4 (1-335)

QY 1 ATGTTATCAATTAAACCTGATGATAGAGTATATCAACCGTTGAAAGCAATA 60
|||||

Db 1 MetValIIeGlnLeuThrProAspAspArgSercIlyTyTyrProProValGluIySgInIle 20
Qy 61 GCAGAGATATATAGTACGTATCTAACTTTAAGCAACAGATGAGGTCATACAGCATCA 120
Db 21 AAlaGlyAspIleValArgIleLeuAsnPhelYsGlnThrAspGluGlyHisThrIleSer 40
Qy 121 TATGGAATTTGAATATGAGCTAAGAAAATAATTTAGCTTACGCTTTGGCTGTAATGGT 180
Db 41 TyGlyIleGluTyArgAlaIySlySIIeIleLeuAlaTyAlaLeuAlaValSerGly 60
Qy 181 ATTCAATATGATCTTAACTTCCCTGAGCTGACTATTTATAGAAATTAAGACTGCTGAGAG 240
Db 61 IleHisAsnValSerIySleuProAspAspTyTyrIlySasnlYsgInThrAlaGluArg 80
Qy 241 ATTATCAAGATATATATGCTTAATCTTTCATCTGCACTATAGCTGAGAAAATGGTGATCA 300
Db 81 IIeTyGlnGlnIyTyMetSerAsnLeuSerSerAlaIleuLeuGlyGluAsnGlyIyAspGln 100
Qy 301 ATTTCTTAAAGATATGCGCAATGGTTTTTATTAAGAAATGACGTGATTTTGAAGTCAATAT 360
Db 101 IIeSerIyAspMetIleAsnGlyPheTyTyrIySasnlGluLeuAspPheGluGlyGlnTy 120
Qy 361 CCTCAAAACATTTGGAATGTTCTGAGCTGGAATAATTAACCATTTGAGTGGCTTATTCAGAT 420
Db 121 ProGlnAsnIIeTyAsnValIProGluIleuGlnIySnlYsProIeuserAlaTyTyrSerAsp 140
Qy 421 GACGATTAATTTATTAAGCACTATATTTTCTCTGTCACAGAAATTCACATGAGAGAAAT 480
Db 141 AspAspIyIleuLeuAlaIeutyTyrPhePheSerValGlnGlnIleProIeuglnGluIyAsn 160
Qy 481 CAACATCAAAATGCCCAAGATTTTATTAATTAATTAATTTCTTATTAACCTTATCTGCT 540
Db 161 GlnGlnSerAsnAlaAlaArgPhePheIyLeuIleAspPheIyLeuPheThrIleuSerAla 180
Qy 541 GTAATCTCACTGGGAGAGAGATTTTTCAAAAAACTTTTACATGATGATAGAGGCTAA 600
Db 181 ValThrSerIeucIyArgArgIIePheSerIySasnlYsPheTyAsnGlyIleuGlnAlaIyS 200
Qy 601 TCATTAGAGATATATATGAGAGAAAAAACTTTCTTAACTTTCTTTCGACCAACCGCAG 660
Db 201 SerIeuglnIyAsnTyTyrIleGlnArgIySlyIeuserIySProPheAspProIeogln 220
Qy 661 AGATTACCTGATGGCAATAGGTTATTTGGCTGGACCAACAGAAACGCTTAATAGAA 720
Db 221 ArgIeProAspIyArgIleGlyTyTyrLeuAlaGlyProThrIyGlnIyAspPolYsTPadY 240
Qy 721 GTGAGTTTAAAGAACTTAAATAATCAAAATCTAGGAATGATTTTCTAATATGAAAGG 780
Db 241 ValSerPheIySglIyIleuTySasnlYsAsnIySserIyArgAsnIyPheSerAsnMetGluGly 260
Qy 781 GCTGCAAAAACAAAGTATAGTTCATTATTAAGAGGTACAAAAGGTTAAGCTCCACAG 840
Db 261 AAlaAlaIySglnIySlyTyrSerSerPheIleIySglnValGlnIySglnAsnAlaProGln 280
Qy 841 ACAGCAACCAAAAGTATGTTGTAACGCCAGTGGCAATGCTGAAAAATTTGCCGAATAT 900
Db 281 ThrAlaAlaIySserIleGlyThrAlaSerGlySerAsnIleuGlnIySleuProAsnAsp 300
Qy 901 TTAATATAGTGTGAGGCTTAAGCAAAAACAGAGGTTAATCTTACTCAAAATGATATGATGAC 960
Db 301 LeuTySerValArgIeuserGlnIySAspArgValThrPheThrGlnIyAsnAspIlyHisP 320
Qy 961 AATACAATGACGGTTCAATAGTGTGGAATCTCATTAATAATAATA 1005
Db 321 AsnThrMetThrValHisSerValGlyThrHisIyTyIySasnlle 335

RESULT 2
US-09-463-048A-3
; Sequence 3, Application US/09463048A
; Patent No. 6630619
; GENERAL INFORMATION:
; APPLICANT: Commonwealth Scientific and Industrial Research Organisation
; APPLICANT: EAST, Peter David

; TITLE OF INVENTION: Toxin Genes from the Bacteria Xenorhabdus nematophilus and Photoc
; FILE REFERENCE: Luminescens
; CURRENT APPLICATION NUMBER: US/09/463,048A
; CURRENT FILING DATE: 2002-12-13
; PRIOR APPLICATION NUMBER: PCT/AU98/00562
; PRIOR FILING DATE: 1998-07-17
; PRIOR APPLICATION NUMBER: PO 8088
; PRIOR FILING DATE: 1997-07-17
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 3
; LENGTH: 368
; TYPE: ERT
; ORGANISM: Xenorhabdus nematophilus
US-09-463-048A-3
Alignment Scores:
Pred. No.: 5,13e-126 Length: 368
Score: 1276.00 Matches: 251
Percent Similarity: 80.17% Conservative: 36
Best Local Similarity: 70.11% Mismatches: 24
Query Match: 73.54% Indels: 24
DB: 4 Gaps: 3
US-10-617-962-2 (1-1008) x US-09-463-048A-3 (1-368)
Qy 4 GTTATACATTAATAACCTGATGATAGA-----AGTGAATATCCACCC 45
Db 11 ValIIeGlnLeuThrProAspAspArgValThrProAspAspIyGlyIyTyArgInPro 30
Qy 46 GTTGAAGCAAAATGACAGAGATATAGTATATCTAACTTTAAGCAACAGATGAG 105
Db 31 ValGluIySglnIleAlaIySIIeIleArgValIleuGlnIySglnIyHisnGln 50
Qy 106 GGTATACAGCATATGGAATTTGAATTTGAGCTGAAGAAAATAATATAGCTTACGCT 165
Db 51 SerHisThrGlyLeuTyTyrIleAlaTyArgAlaIySlyValIleIleAlaTyArgAla 70
Qy 166 TTGGCTGATAGTGTATTCATATGATCTTAACTTTCGATGACTATTAATAGATAAA 225
Db 71 LeuAlaValSerGlyIleHisAsnValSerGlnIySleuProGluAspIyTyTyrIySasnlYs 90
Qy 226 GAGACTGCTGAAGAGATTTATCAAGATATATGCTAATCTTTCACTGCACTATTAAGT 285
Db 91 AspAsnThrGlyArgIleTyGlnGlnIyTyMetSerAsnLeuIleuSerIleAlaIySgln 110
Qy 286 GAAATATGATCAAAATTTCTAAGATATGCAAAATGTTTATTAAGAAATGAATGAT 345
Db 111 GluAsnGlyAspGlnIleSerIySasPheAlaAsnAspPheThrGlnAsnGluIySgln 130
Qy 346 TTTGAAGTCAATATCTCAAAACATTTGGAATGTTCTGAGCTTGAATAATTAACCATTTG 405
Db 131 PheGlyGlyGlnArgIeIySasnlYsAsnThrIyAspIleProAspIeIySglnIySleuIyS 150
Qy 406 AGTCTTATTCAGATGACGATTAATTTATAGCACTATATTTTCTCTGTAACGAAAT 465
Db 151 GluAspTySerAspIySAspIySleuIleuAlaIeutyTyrPheAspAlaSerGlnIySgln 170
Qy 466 CCACTGAGAGAAATCAAAATCAAAATGCCGAAGATTTTAAATTAATTAATGATTTCTTA 525
Db 171 PrometGluAlaAsnIySglnIySAsnAlaAlaAsnAspPheIySValIleAspPheIyS 190
Qy 526 TTTACCTTATTCGCTGTAATCTCACTGAGAGAGATTTTTCAAAAAACTTTTACAT 585
Db 191 LeuIleuSerAlaValThrSerIeucIyTySgIIePheSerIySasnlYsAsnIySAsn 210
Qy 586 GATTTAGAGCGTAAATCATTAAGATTTATTTAGAGAGAAAAAACTTTTCAACCTTTTC 645
Db 211 GlyIleuGlnIyThrIySerIeucIySAsnTyTyrIleGlnArgIySlyIeuserIySProPhe 230
Qy 646 TTTGACCAACCGCAAGATTAATCTGATGAGCAAAATAGGTTATTTGGCTGCAACAGAA 705


```

Db      231 pheargProProGlnInlyLeuProAspGlyArgThrGlyTyrLeuAlaGlyProThrIlys 250
QY      706 GGGCCATAA----- 714
Db      251 AlaProLyLeuProThrThrsSerSerThrAlaThrThrsSerThrAlaAlaSerSerAsn 270
QY      715 TCGAGAGTGAAGTTTAAAGACTTAAATAAATAAATACTAGAAATGATTTTCTAATATG 774
Db      271 TPAAGValSerLeuGlnInlyLeuAlaArgAsnProSerArgAsnThrhehelelySmet 290
QY      775 GAAGGGCTGCAGAAACAAAGTATAGTTCATTATTAAGAGGTACAAAGGGTAAAGCT 834
Db      291 AspaRPaLaAlaLyAsArgLySerSerSerPheLeIleGlyValGlnInlySGIlyAsnAsp 310
QY      835 CCAACAGACGACGCGAAAGTATGTTGTAAGCCAGCGAGTAACTGGAATAATTTGCCG 894
Db      311 ProAlaGAlaAlaAlaAlaSerIleGlyThrLySerGlySerAsnheGlnIlyLeuGln 330
QY      895 ---AATTAATTTATATAGTGTGAGGCTAAGCCAAAGAGAGGTAACCTTTACTCAAAAT 951
Db      331 GlyArgAspLeuTySerIleArgLeuSerGlnGlnIleAlaArgValThrheSerIleAsn 350
QY      952 GATACTGACAAATACAAATGACGGTTCATAGCTGTGAACTCATTTATAAATAATA 1005
Db      351 AsenThrAspGlnIleMetGluIleGlnSerValGlyThrIleTyGlnAsnIle 368

RESULT 3
US-08-569-168-7
; Sequence 7, Application US/08569168
; Patent No. 5972687
; GENERAL INFORMATION:
; APPLICANT: SmigielSKI, Adam J.
; TITLE OF INVENTION: TOXIN GENE FROM XENORHABDUS NEMATOPHILUS
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lowe, Price, Leblanc & Becker
; STREET: 99 Canal Center Plaza, Suite 300
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22314
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/569,168
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Price, Robert L.
; REGISTRATION NUMBER: 22,685
; REFERENCE/DOCKET NUMBER: 1451-015
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-684-1111
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 383 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-569-168-7

Alignment Scores:
Pred. No.: 5,06e-103
Score: 1058.00
Percent Similarity: 73.688
Best Local Similarity: 65.108
Query Match: 60.988
DB: 2
Length: 383
Matches: 235
Conservative: 31
Mismatch: 58
Indels: 38
Gaps: 10

```

```

US-10-617-962-2 (1-1008) x US-08-569-168-7 (1-383)
QY      4 GTTATACAAATTAAACCCGATGATGA-----AGTGATATGCCACC 45
Db      11 ValIleGlnLeuThrProAspAspArgValThrProAspAspArgGlyGluIlyGlnPro 30
QY      46 GTTGAAGAACAAATAGCAGAGATATAGTACTATTAACCTTAACCAACAGATGAG 105
Db      31 ValGluIySGlnIleAlaGlyAspIleIleArgValIleuGlnPheIySGlnThrheGln 50
QY      106 GGTATACAGCATCATATGAAATGAAATGAGGCTAAAGAAATAATATTAAGTACGCT 165
Db      51 SerIleThrGlyLeuTyGlyIleProIyArgAlaIlyLeuValIleIleAlaTyAla 70
QY      166 TTGGCTGTAGAGGATTCATTAATGATTAATTAACCTTCTGAGTATTAATGAATAA 225
Db      71 LeuAlaValSerGlyIleHisAsnValSerGlnLeuProGluAspIlyTyIlyAsnIly 90
QY      226 GAGACTGCTGAGAGAAATTTATCAAGATATATGCTTAATCTTCACTGACATTAAGT 285
Db      91 AspaenThrGlyArgIleTyGlnValIlyMetSerAsnLeuSerAlaLeuLeuGly 110
QY      286 GAAATGTGTATCAAAATTTCTAAGATATGCGAAATGCTTTTATAGAAATGAGTGAAT 345
Db      111 GluAsnGlyAspGlnIleSerIyAspMetAlaAsnAspPheThrGlnAsnGlyLeuGln 130
QY      346 TTTGAGAGTCAATATCTTCAAAACATTTGGAATGCTTCTGAGCTGAAATTAACCATG 405
Db      131 PheGlnI-ValAsnValIleuIySileProGlyIlePheLeuIleuAspGlyIleAsnIlyTr 150
QY      406 -AGTGCTTATTCAGATGACGATTAATTTATAGCATATATTTTCTGCTGACGAGAAAT 464
Db      150 PLYSileTySerAspGluAspLyLeuLeuAlaLeuTyRhePheAlaSerGlnGlyIle 170
QY      465 TCCACTGAGAGAAATCAACATCAATGACCGCAAGATTTTAAATTAATGATTTCTT 524
Db      170 UprometGluAlaAsnGlnInSerAsnAlaAlaAsnPheIyValIleAspPheIle 190
QY      525 ATTTCCTTATCTGCTGATCTTCACTGGAAGAGATTTTTCAAAAAATTTCACAA 584
Db      190 UleuIleuSerAlaValAlThrSerLeuGlyIyAspGlyIlePheSerIyAsnPheTyAs 210
QY      585 TCGATTAGAGGCTAATCATTAAGAAATTAATTAAGAAATAAATCTTCAACCTT 644
Db      210 neIyLeuGlnThrLySerLeuGlnAsnIlyIleGluAspLyIyLeuSerIyPheProPh 230
QY      645 CTTTGCACCAACCGCAGAGATTAAGTATGCGAAGATTAATTTGCTGACCAACAGA 704
Db      230 ePheArgProProGlnIyLeuProAspGlyArgThrGlyTyrLeuAlaGlyProThrIly 250
QY      705 AGCGCTTAA----- 714
Db      250 salAProIyLeuProThrThrsSerSerThrAlaThrThrsSerThrAlaAlaSerSerAs 270
QY      715 -TGCAGAGTGAAGTTTAAAGAACTTAAATAAATAAATCTAGANATTTTCTAATAT 773
Db      270 nTPArgValSer-----LeuGlnIySerProGlnIleGlnIyTyIle--Tyr 286
QY      774 GGAAGGGGC---TGCAGAAACAAAGATATAGTTCATTATTAAGAGGTACAAAGGGTAA 830
Db      287 GluAsnGlyCyCyLeuThrIlySile---PheIleTyIyLeuArgGlyIlyThrIySgly 304
QY      831 CGCTCCACAGACAGACGCGAAAGTATGTTGTAAGCCAGCGAGTCACTGAAATAAT 890
Db      305 ---SerThrCySerSerSerIySerTyIlyTyIyLySArgGln--LeuArgIyThr 322
QY      891 GCCGATATATTATATAGTGAAGGCTAAGCCAAAGAGAGGTAACCTTATCAATAA 950
Db      323 AlaArgAspPheIle---TyIyThrIySProArgIyGlnIyAsnIleLeuIlySlys 341
QY      951 TGATCTGACAAATACAAATGACGTTTCATAGTGTGAACCTCATTAATAAATAATATAG 1007

```

Db 342 ---TyrProAsnAsnGlyAspProlys---CysTrpAsnSerLeuProLysTyrIle 358

RESULT 4
US-08-569-168-2
Sequence 2, Application US/08569168
Patent No. 5972687
GENERAL INFORMATION:
APPLICANT: Smigielaki, Adam J.
APPLICANT: Akhurst, Raymond J.
TITLE OF INVENTION: TOXIN GENE FROM XENORHABDUS NEMATOPHILUS
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSER: Lowe, Price, Leblanc & Becker
STREET: 99 Canal Center Plaza, Suite 300
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22314
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/569,168
FILING DATE:
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Price, Robert L.
REGISTRATION NUMBER: 22,685
REFERENCE/DOCKET NUMBER: 1451-015
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-684-1111
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 278 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-569-168-2

Alignment Scores:
Pred. No.: 1,11e-90 Length: 278
Score: 940.50 Matches: 191
Percent Similarity: 80.22% Conservative: 24
Best Local Similarity: 71.27% Mismatches: 29
Query Match: 54.21% Indels: 25
DB: 2 Gaps: 2

US-10-617-962-2 (1-1008) x US-08-569-168-2 (1-278)

QY 4 GTTATACAAATTAACACCTGATGATAGA-----AGTGATATCCACCC 45
|||
Db 11 ValIleGlnLeuThrProAspAspArgValThrProAspAspLysGlyValGlyArgGlnPro 30
|||
QY 46 GTTGAAGAACAATATACAGAGATATAGTACGTATCTTAACCTTAAAGCAACAGATGAG 105
|||
Db 31 ValGlnLysGlnIleAlaGlyAspIleIleArgValLeuGlnIlePheLysGlnThrArgGln 50
|||
QY 106 GGTCAATACAGCATATATGAAATTTGAAATTCGAGCTTAAGAAAATAATATATGCTTACGCT 165
|||
Db 51 SerHisThrGlyLeuValGlyLeuProTyrArgAlaLysLysValIleIleAlaTyrAla 70
|||
QY 166 TTGGCTGTAGTGTATGATATGATATGATTAACCTTCCGATGACTTATTATAGAAATAA 225
|||
Db 71 LeuAlaValSerGlyIleHisAsnValSerGlnLeuProGlnAspLysTyrIleLysAsnLys 90
|||
QY 226 GAGACTGTGAGAGAAATTTATCAAGATATATATGCTTAATCTTCATCTGCACCTATTAGGT 285
|||
Db 91 AspAsnThrGlyArgIleTyrGlnValTyrMetSerAsnLeuSerAlaLeuLysGly 110
|||
QY 286 GAAATGTGATGCAAAATTTCTAAAGATATGCGAAATGCTTTTATATAGAAATGAACTGAGAT 345
|||

Db 111 GluAsnGlyAspGlnIleSerLysAspMetAlaAsnAspPheThrGlnAsnGlnLeuGlu 130
|||
QY 346 TTTGAAGGTCAATATCTTCAAAAACATTTGGAATGTTCTGACCTTGAAATAAACCATTT 405
|||
Db 131 PheGln-ValAsnValLeuValIleProGlyIlePheLeuIleLeuArgIleAsnTyrTr 150
|||
QY 406 -AGTGCTTATTCAGATGACGATTAATTTATTTAGCATTATTTTCTCTGTACAGGAAT 464
|||
Db 150 PheLysIleTyrSerAspGlnAspLysLeuLeuValLeuTyrPhePheAlaSerGlnGlu 170
|||
QY 465 TCCACTGAGGAGAAATCAACATCAAAATGCGCAGATTTTAAATTAATGATTTCTT 524
|||
Db 170 uPheMetGlnAlaAsnGlnSerAsnAlaAlaAsnPheLysValIleAspPhe 190
|||
QY 525 ATTACCTTATCTGCTGATCACTTCACTGGAGAGAGATTTTTCAAAAAATTATACAA 584
|||
Db 190 uLeuIleLeuSerAlaValThrSerLeuGlyAspArgIlePheSerLysAsnPheTyrAs 210
|||
QY 585 TCGATTAGAGCTTAATCATTTAGAGATTAATTATAGAGAAAAAATTCTTAACCTTT 644
|||
Db 210 nGlyLeuGlnThrLysSerLeuGlnAsnTyrIleGlnArgLysLysLeuSerLysPhePro 230
|||
QY 645 CTTTGACACCGCAGAGATTAAGTCTGATGCGAGAAATAGGTTATTTGGCTGACCAACAG 704
|||
Db 230 ePheArgProProGlnLysLeuProAspLysArgThrGlyLysLeuAlaGlyProThrLys 250
|||
QY 705 AGCGCCTAAA----- 714
|||
Db 250 sAlaProLysLeuProThrThrSerSerThrAlaThrThrSerThrAlaSerSerAs 270
|||
QY 715 -TGAGAGTGAAGTTTAAAGAA 735
|||
Db 270 nTrpArgValSerLeuGlnLys 277
|||

RESULT 5
US-09-328-352-6677
Sequence 6677, Application US/09328352
Patent No. 6562958
GENERAL INFORMATION:
APPLICANT: Gary L. Breton et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
FILE REFERENCE: GTC99-03PA
CURRENT APPLICATION NUMBER: US/09/328,352
CURRENT FILING DATE: 1999-06-04
NUMBER OF SEQ ID NOS: 8252
SEQ ID NO 6677
LENGTH: 816
TYPE: PRT
ORGANISM: Acinetobacter baumannii
US-09-328-352-6677

Alignment Scores:
Pred. No.: 0.0184 Length: 816
Score: 105.50 Matches: 96
Percent Similarity: 30.29% Conservative: 40
Best Local Similarity: 21.38% Mismatches: 128
Query Match: 6.08% Indels: 185
DB: 4 Gaps: 22

US-10-617-962-2 (1-1008) x US-09-328-352-6677 (1-816)

QY 112 ACAGCATCATATGATGATTAATGAGCTAAGAAATTAATATAGCT-----TAC 162
|||
Db 216 ThrSerAlaGlnAsnGlnThrTyrThrValAlaGlyValIleThrAlaAspTyrArgTyr 235
|||
QY 163 GCTTGGCTGTAGTGTATTCATATGATTAATGATTAACCTTCCGATGACTATTATAGAAT 222
|||
Db 236 AlaAsnGlyPheSerGlyPheTyr-----ValGlnThrProAspThr-----LysAla 251
|||
QY 223 AAGAGACTGTGAGAGAAATTTATCAAGATATATATGCTTAATCTTCATCTGCACATATA 282
|||

```

Db      252 ArgAlaAsnValSerAsnAlaIlePheValIYrIleProAsnSerSerAlaValIYsGly 271
QY      283 GGTGAAGAAATGATGAAATTTCTAAGATATGCGAAATGCTTTTATAGAAATGACG 342
Db      272 GlyGlnIleGlyAspGlnValIleLeuArgIYArgLeuThrSerTYrGlnAsnGlnLeu 291
QY      343 GATTTTGAAGTCATATCTCAAAAACATTTGGAAATGT---CGTGAAGCTTGAAT--- 396
Db      292 GlnIleAsp---GlnLeuGlnGlnAspIleGlnThrCysAsnSerSerMetAlaAsnGln 310
QY      397 ---AAACCATTAAGTGGCTTATTTCAGATGACGATAAATTAATGACCTAATATTTTCTCT 453
Db      311 ValGlnProIleSer-----LeuGlnLeuProPheSerSer 322
QY      454 GTA----- 456
Db      323 LeuThrAlaGlySerThrHisSerProGlnArgTYrGlnGlyMetLeuValIYsLeuPro 342
QY      457 CAGAAATTCACCTGAGAGAAATCAACATCAATCAATGCGCAAGATTTTAAATTAAT 516
Db      343 GlnThrLeuThrValSerGlnAsnTYr---AsnTYrGlyArgTYrGlyGlnLeu--- 359
QY      517 GATTTCTTAATTAACCTTAATCTGCTGTAACCTGCACTGGAGAGAGATTTTCAAAAAAC 576
Db      360 -----SerLeuSerLeuGlyArgLeuTYrIleProThrAsn 371
QY      577 TTTTCAATGATTA-----GAGGCTAATCATTGAGAAATTAATTAAGAGAAAAA 630
Db      372 LeuTYrProAlaLeuSerProGlnAlaValAlaLeu-----AlaGlnIYsAsnLeu 388
QY      631 CTTTCTAAACCTTTCTTT-----CGACCAACCGCAGAGATTA 666
Db      389 LeuSerLeuIleIlePheAspAspGlyTYrAsnAsnGlnAsnArgThrProTrp--Leu 407
QY      667 CCTGATGCG-----AGATAGATTATTTGGCTGACCAACAGAA 705
Db      408 ProThrAsnPheSerValAlaAsnThrLeuArgSerGlyTYrGlnLeuIYsAsnValGln 427
QY      706 GCG-----CCTAAATGAGAGATG--- 723
Db      428 GlyIleLeuGlnTYrArgPheAsnGlyTYrPArgValGlnProValLeuGlyArgThrGln 447
QY      723 ----- 723
Db      448 ProGlnValIleThrGlnThrAsnProAlaGlnAsnIleIleThrIYsAsnAlaAsnHis 467
QY      724 -----AGTTTAAAGAACTTAAATAATACAAATCTAGAAATGATTTCTAAT 771
Db      468 IleArgValAlaSerPheAsnValIleuAsnTYrAspAsnGlyAlaThrGlyPheProThr 487
QY      772 ATGGAAGGGCTGCAGAAACAA-----AAGTATAGTTCAATT 807
Db      488 GlnArgGlyAlaAsnThrGlnAlaGlnPheAspIYsGlnHisIYsIleValSerAla 507
QY      808 ATAAAAGAGTA-----CAAAAGGCTAACGCT 834
Db      508 LeuIYsSerIleAspAlaAspValTYrGlyLeuMetGlnIleAlaAsnAsnGlyTYrGly 527
QY      835 CCACAGACAGCAGCAAAAGATTTGGTACACCGCAGTGC----- 873
Db      528 ProAsnSerAlaIleAlaHisLeuThrSerAlaLeuGlyIProAspTYrIYsTYrValIle 547
QY      874 ---AGTAACTGGAATAATTTCCGATTAATTAATAGTGTGAG--- 915
Db      548 ProGlnAsnLeuAspArgLeuGlyAsnAspValIleAlaValAlaIleIleTYrAsnSer 567
QY      916 -----CTAAGCAAAAGAGAGAGTA 936
Db      568 IYsArgValIYsProLeuAsnIYsAlaValValLeuAspLeuGlyAspIYsAsnArgThr 587
QY      937 ACCTTACTCAA----- 948
Db      588 ThrLeuAlaGlnThrPheGlnAlaValArgGlyAsnIYsIlePheThrValIleProAsn 607

```

```

QY      949 -----AATGATGACATACAA 966
Db      608 HisLeuIYsSerIYsGlyCysSerGlyValAlaPheIaSerSerSerAspAlaAspGln--A 627
QY      967 ATGACGGTTCATATGTTGGAAAC 989
Db      627 AsnAspGlyGlnGlyCysTrpAsn 634

RESULT 6
US-09-248-796A-23971
; Sequence 23971, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstein et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; PRIOR FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 23971
; LENGTH: 600
; TYPE: PRT
; ORGANISM: Candida albicans
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (590),(595),(597)
; OTHER INFORMATION: Identity of amino acid sequences at the above locations are unkno
US-09-248-796A-23971

Alignment Scores:
Pred. No.: 0.0435 Length: 600
Score: 101.50 Matches: 62
Percent Similarity: 39.56% Conservative: 46
Best Local Similarity: 22.71% Mismatches: 110
Query Match: 5.85% Indels: 55
DB: 4 Gaps: 13

US-10-617-962-2 (1-1008) x US-09-248-796A-23971 (1-600)
QY      202 CCTGATGACTATTAATGAC-----AATAAAGAGACTGCTGAGACA 240
Db      286 ProPheSerPheTYrGlnSerAlaSerAlaAsnSerSerAsnSerSerLeuIYsThrIYs 305
QY      241 ATTTATCAAGATATATATGCTAATCTTTCATCTGACATTA-----GGTGAAT 291
Db      306 ThrIleGlnIYsThrLeuSerAsnSerSerSerSerSerLeuPheAlaThrProArgSer 325
QY      292 GGTGATCAAAATTTCTAAAGATATGCAATGCTTTTATTAAGATGACCTGATTTTGA 351
Db      326 GlyAlaIYsAlaSerArg-----TYrIYsAsnIleTYrAsp----- 337
QY      352 GGTCAATATCTCAAAACATTTGGAATGTTCTCTGACTTGAATAAACCATTGAGTCT 411
Db      338 -----GlnSerIYsIleIleSerAsnAlaMetIYsIleSerSerGlnAsnValSerGly 355
QY      412 TATTCAGATGACGATTAATTAATTAAGACATATTTTCTCTGACAGAAATTCACATG 471
Db      356 IleGln-----ArgThValThrProProLeu 364
QY      472 GAGAAATCAACAAATCAATCAATGCGCAGATTTTAAATTAATGATTTCTTAATTAC 531
Db      365 AsnArgAsnAsnAsnAsnAsnSerAsn-----LYsTYrIleuAsnHis 379
QY      532 TTAATGCTGCTGATCTTCACTG---GGAAGAGATTTTTCAAAAACCTTTACATGGA 588
Db      380 SerSerSerLeuProSerIleuIYsGlyIYsArgSerPheSer----- 393

```

```

QY 589 TTAGAGGCTAAATCATTAGAGATTATATTGAGAGAAAAAACTTTCTAAACCT----- 642
DB 394 ---TyrValArgGlnLeu---ArgTyrGlyAspSerLysArgValAlaSerProLysThr 411
QY 643 -----TTCTTTCGACCGCGAGATTACCTGATGGCAGATAGCTTATTGGCTGGA 696
DB 412 AenIleThrIleAspThrProPheThrProPheThrAlaThrAlaThrAlaThrAlaThr 431
QY 697 CCACAGAGACGCTAAATGAGAGTGTGATTGAACTTAAATAACAATCTAG 756
DB 432 AlathThrValProThr---SerValSerAlaValThrValLeuSerSerIleGlyLeu 450
QY 757 AATGATTTTCTAATATGAGAGGGCTGCAGAAAACAAGTATGCTCATTTATTAAGAG 816
DB 451 SerGlyLys-----SerProSerSerLysSerProGlnThrGlnPheIleLeuGln 466
QY 817 GTACAAAGAGGTAAACGCTCACAGACAGCGAGAAAGTATGGTACGCGAGTGCAGT 876
DB 467 LeuProSerAlaSerThrProSerThrProSerThrSerAlaThrThrLeuSerThrSerAlaSer 486
QY 877 AACCTGAGAAATTTGCCGATATTTATATTAGTGTGAGGCTTAAAGCCAAAAGACAGGCTA 936
DB 487 GlnLeuLysLysAsnArgAsnSerProValSerValAsnSerGlnAlaLeuAspSerTyr 506
QY 937 ACCTTACTCAAAATGATATCTGACAAATACATGACGCTT 975
DB 507 Thr---ThrSerSerSerThrAsnSerThrMetAspIle 518

```

RESULT 7

```

US-08-628-291-4
; Sequence 4, Application US/08628291
; Patent No. 5801031
; GENERAL INFORMATION:
; APPLICANT: GALIVAN, JOHN H.
; APPLICANT: RYAN, THOMAS J.
; APPLICANT: YAO, RONG
; APPLICANT: NIMEC, ZENIA
; TITLE OF INVENTION: GAMMA GLUTAMYL HYDROLASE
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON, HARGRAVE, DEVANS & DOYLE LLP
; COUNTRY: USA
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/628,291
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: TIMIAN, SUSAN J.
; REGISTRATION NUMBER: 34,103
; REFERENCE/DOCKET NUMBER: 20894/150
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 716-263-1636
; TELEFAX: 716-263-1600
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 293 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-628-291-4

```

Alignment Scores:

```

Pred. No.: 0.113 Length: 293
Score: 96.50 Matches: 51
Percent Similarity: 40.57% Conservative: 35
Best Local Similarity: 24.06% Mismatches: 99

```

```

Query Match: 5.56% Indels: 27
DB: 1 Gaps: 11
US-10-617-962-2 (1-1008) x US-08-628-291-4 (1-293)
QY 244 TTTCAGAAATATATGCTTAATCTTCATCTGACATTAATGAGAAATGCTGATCAATT 303
DB 58 TyrGlnThrLeuPheArgSerIleAsnGlyValLeuLeuProGlyGlyValAsnLeu 77
QY 304 TCTAAGAT-----ATGCAATAGGTTTATTAAGAAATGACCTGATTTT--- 348
DB 78 ThrIleSerGlyTyrSerArgValAlaLysIlePhePheThrLysAlaLeuIleSerPhe 97
QY 349 ---GAAGTCAATAT---CTCAAAAACATTGGAATGTTCTGACCTGAGAAATTAACCA 402
DB 98 AspAsnGlyAspTyrPhePro-----ValTrpGlyThrCys---LeuGlyLeuIleGln 114
QY 403 TTGAGTGTCTTATTCAGATGACGATAAATTAATAGCATTAATTTTCTCTGACAGAA 462
DB 115 LeuSerValLeuValSerAsnAspAsnLeuThrLeuThrAsnThrSerSerValLys 134
QY 463 ATTCCACTGAGAGAAATCAACAATCAATCAATGCGCAGAGATTTTAAATTAATGATTTC 522
DB 135 LeuProLeuAsnPheThrArgAspSerLysGlnSerArgMetPheArgAsnLeu----- 152
QY 523 TTATTTACCTTATCTGCTGTAATTCTACTGGAGAGAGATTTTCAAAAACCTTTTAC 582
DB 153 -----ProGlnGlnLeuLeuAsnSerLeuAlaSerGlnAsnLeuThrAlaAsnPheHis 170
QY 583 AATGATTAAGAGGCTTAATCATTAGAGAAATTATATTGAGAGAAAAAACTTTCTAAACCT 642
DB 171 LysTrp-----SerLeuSerValLysAsnPheThrGlnAsnGlnLysLeuLys--- 187
QY 643 TTCTTTCGACACCCAGAGATTACCTGATGCGCAGAAATAGTTATTGGCTGGA----- 696
DB 188 PhePheAsnIleLeuThrValAsnThrAspGlyLysThrGlnPheIleSerSerMetGlu 207
QY 697 -----CCACAGAGCGCTTAATGAGAGTGTGATTGAAAGCTTAAAGAACTTAATAAT 744
DB 208 GlyTyrLysTyrProIleTyrAlaValGlnTrpHisProGlnLysAlaProPheGlnTrp 227
QY 745 AACCAATCTAGAAATGATTTTCTTAATATGAGAGGGCTGCAGAAACAAAGAT----- 798
DB 228 LysLysLeuArg---GlyLysSerHisAlaProAsnAlaValLysThrSerPheTyrLeu 246
QY 799 ---AGTTCAATTATTAAGAGGTACCAAAAGGCTAAC 831
DB 247 AlaLysPhePheIleSerGlnAlaLeuLysAsnAsp 258

```

RESULT 8

```

US-09-128-722-4
; Sequence 4, Application US/09128722
; Patent No. 5962235
; GENERAL INFORMATION:
; APPLICANT: Galivan, John H
; APPLICANT: Ryan, Thomas J
; APPLICANT: Yao, Rong
; APPLICANT: Nimec, Zenia
; TITLE OF INVENTION: Gamma Glutamyl Hydrolase
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Jaecle Fleischmann & Muegel, LLP
; STREET: 39 State Street
; CITY: Rochester
; STATE: New York
; COUNTRY: US
; ZIP: 14614-1310
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:

```

```

: APPLICATION NUMBER: US/09/128, 722
: FILING DATE: 04-AUG-1998
: CLASSIFICATION: 514
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/628,291
: FILING DATE: 05-APR-1996
: ATTORNEY/AGENT INFORMATION:
: NAME: Braman, Susan J
: REGISTRATION NUMBER: 34,103
: REFERENCE/DOCKET NUMBER: 87681,98R196
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 716-262-3640
: TELEFAX: 716-262-4133
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 293 amino acids
: TYPE: amino acid
: STRANDEDNESS: not relevant
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-09-128-722-4

Alignment Scores:
Pred. No.: 0.113 Length: 293
Score: 96.50 Matches: 51
Percent Similarity: 40.57% Conservative: 35
Best Local Similarity: 24.06% Mismatches: 99
Query Match: 5.56% Indels: 27
DB: 2 Gaps: 11

US-10-617-962-2 (1-1008) x US-09-128-722-4 (1-293)
QY 244 TATCAAGATATATATGCTTAATCTTTCATCTGCACCTATTAGGAAAAATGGTGATCAAT 303
DB |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
58 TgTgUthtLeuPhenArgSerIleAenGlyValLeuLeuProGlyGlyAlaLeuLeu 77
QY 304 TCTAAAGAT-----ATGCGAATGGCTTTTATTAAGATGACTCGATTT-- 348
DB ::::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
78 ThrHisSerGlyTyrSerArgValAlaIlePhePheThrLysAlaLeuGluSerPhe 97
QY 349 ---GAAGTCATAT---CCTCAAAACATTGGAAATGTTCTGAGCTTGAATAATACGA 402
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
98 AspaenGlyAspTyrPhePro-----ValTpgGlyThrCys--LeuGlyLeuGlu 114
QY 403 TTGAGTGCTTATTCAGATGACGATTAATATTATGACATATATTTTCTGTACAGAA 462
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
115 LeuSerValLeuValSerAsnAspAsnLeuLeuThrLeuThrAsnThrSerSerValLys 134
QY 463 ATTCACCTGGAGAAAAATCAACAATCAAAATCCGCAAGATTTTAAATTAATTGATTC 522
DB ::::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
135 LeuProLeuAsnPheThrArgAspSerLysGlnSerArgPheArgAsnLeu----- 152
QY 523 TTATTACTTATCTGCTGTAACTTACCTGCGGAAGAGATTTTTCAAAAAATTTC 582
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
153 -----ProGluGluLeuLeuAsnSerLeuAlaSerGluLeuThrAlaAsnPheHis 170
QY 583 AATGATTTAGAGGCTAAATCATTAAGAAATTTATTTAGAGAAAAAACTTTTAACT 642
DB ::::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
171 LysTrp-----SerLeuSerValLysAsnPheThrGlnAsnGlyLysLeuLysLys--- 187
QY 643 TTCCTTCGACCAACCGCAGAGATTACCTGATGCGAAATAGGTTATTTGGCTGA----- 696
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
188 PhePheAsnIleLeuThrValaLeuThrAsnArgGlyLysThrGluPheThrIleSerSerMetGlu 207
QY 697 -----CCAACAAGAGGCTTAATGAGAGTGAAGTGTATTAAGAATTAAAT 744
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
208 GlyTyrLysTyrProLetyrAlaValGlnTrpHisProGluLysAlaProPheGluTrp 227
QY 745 AACCAATTCAGAAATGCTTTCTAATATGGAAGGCGCTCAAAACAAAGTAT----- 798
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
228 LysLysLeuArg---GlyIleSerHisAlaProAsnAlaValLysThrSerPheTyrLeu 246
QY 799 ---AGTTCATTTATTAAGAAGGTACAAACGGGTAC 831

```

```

DB      247 AlalysPhephelleSerGlnLalaLeuYsaAaNap 258
RESULT 9
US-08-628-291-12
Sequence 12, Application US/08628291
Patent No. 5801031
GENERAL INFORMATION:
APPLICANT: CALIVAN, JOHN H.
APPLICANT: RYAN, THOMAS J.
APPLICANT: YAO, RONG
APPLICANT: NIMEC, ZENIA
TITLE OF INVENTION: GAMMA GLUTAMYL HYDROLASE
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON, HARGRAVE, DEVANS & DOYLE LLP
COUNTRY: USA
ZIP: 14603
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/628, 291
FILING DATE:
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: TIMIAN, SUSAN J.
REGISTRATION NUMBER: 34, 103
REFERENCE/DOCKET NUMBER: 20894/150
TELECOMMUNICATION INFORMATION:
TELEPHONE: 716-263-1636
TELEFAX: 716-263-1600
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 317 amino acids
TYPE: amino acid
STRANDNESS: not relevant
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-628-291-12
Alignment Scores:
Pred. No.: 0.116 length: 317
Score: 96.50 Matches: 51
Percent Similarity: 40.57% Conservative: 35
Best Local Similarity: 24.06% Mismatches: 99
Query Match: 5.56% Indels: 27
DB: 1 Gaps: 11
US-10-617-962-2 (1-1008) x US-08-628-291-12 (1-317)
QY      244 TATCAAGATATATATGTCTTAACTTCATCGCATTTAGTGAAAATGGTGATCAAATT 303
      |||::: ::::: |||||
DB      82 TyrgIutrlrleuPheargserileanglyValleuProclgyIglylaAsnleu 101
QY      304 TCTTAAGAT-----ATGCGAATGGTTTTTAAGAAGTAACGCGATTT--- 348
      ::::: ::||| ::::: ||::::
DB      102 ThrhisSergIyTSerArghValAlaIylsilPhepherlrylsAlaleuGluSerPhe 121
QY      349 ---GAAGCATATAT---CCGCAAAACATTGGAAATGGTCTCGAGCTTGAAAAATAACA 402
DB      122 AspaangIyAspTy-PhePro-----ValTrpGIyThCyS---leuGIyleuGIuGlu 138
QY      403 TTGAGTGCTTATTCAGANGACATATAATTATTAGCACATATATTTTTTCTCTGACAGGAA 462
DB      139 leuSerVallleuValSerAsnaSpasndleuThrleuThraSnthrSerSerVallys 158
      ||||| |||::: ::||| ||||| |||||
QY      463 ATTCACACTGGAGAAAATCAACAATCAACAATCGCAAGAGATTTTAAATTAATTGATTTTC 522
DB      159 leuProleuAsnPhethrArgAsberIylsGlnSerArgMeChargAsnleu----- 176

```

```

QY 523 TTATTTACCTTATCTGCTGTAACCTGACGGAGAGATTTTTCAAAACTTTTAC 582
Db 177 -----ProGlugluLeuLeuLeuSerLeuAlaSerGluAsnLeuThrAlaAsnPheHis 194
QY 583 AATGATTAGAGCTAAATCATTAGAAATTATATTGAGAGAAAAAACTTTCTAACT 642
Db 195 LysTP-----SerLeuSerValLysAsnPheThrGluAsnGluLysLysLys--- 211
QY 643 TTCTTTCAGCACCGCAGAGATTAACCTGATGGCAGATAGATTATTTGGCTGGA----- 696
Db 212 PhePheAsnIleLeuThrValAsnThrAspGlyLysThrGluPheIleSerSerMetGlu 231
QY 697 -----CCACAGAGCGCTTAATGAGAGTAGATTTAAGAACTTAATAAAT 744
Db 232 GlyTyrLysTyrProIleTyrAlaValGlnTyrPheProGluLysAlaProPheGluTyr 251
QY 745 AACCAATCAGGAATGATTTTCTATATGAAAGGGGCTGCAAAACAAAGTAT----- 798
Db 252 LysLysLeuArg---GlyIleSerHisAlaProAsnIleValLysThrSerPheTyrLeu 270
QY 799 ---AGTTCATTATTAAGAGGTACAAAGGGTAAAC 831
Db 271 AlaLysPhePheIleSerGluAlaLeuLysAsnAsp 282

RESULT 10
US-09-128-722-12
; Sequence 12, Application US/09128722
; Patent No. 5962235
; GENERAL INFORMATION:
; APPLICANT: Galivan, John H
; APPLICANT: Ryan, Thomas J
; APPLICANT: Yao, Rong
; APPLICANT: Nimec, Zenia
; TITLE OF INVENTION: Gamma Glutamyl Hydrolase
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSES: Jaeckle Fleischmann & Wogel, LLP
; STREET: 39 State Street
; CITY: Rochester
; STATE: New York
; COUNTRY: US
; ZIP: 14614-1310
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/128,722
; FILING DATE: 04-AUG-1998
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/628,291
; FILING DATE: 05-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Braham, Susan J
; REGISTRATION NUMBER: 34,103
; REFERENCE/DOCKET NUMBER: 87681,98R196
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 716-262-3640
; TELEFAX: 716-262-4133
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 317 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-128-722-12

Alignment Scores:
Pred. No.: 0.116 Length: 317
Score: 96.50 Matches: 51

```

```

Percent Similarity: 40.57% Conservative: 35
Best Local Similarity: 24.06% Mismatches: 99
Query Match: 5.56% Indels: 27
DB: 2 Gaps: 11

US-10-617-962-2 (1-1008) x US-09-128-722-12 (1-317)

QY 244 TATCAAGATATATATGCTAAATCTTTCATCTGCACTATTAAGTGAATAATGATCAAT 303
Db 82 TyrGluThrLeuPheArgSerIleAsnGlyValLeuLeuProGlyGlyIleAsnLeu 101
QY 304 TCTAAAGAT-----ATGCCAATAGGTTTATTAAGAAAGAACTGATTTT--- 348
Db 102 ThrHisSerGlyTyrSerArgValAlaLysIlePhePheThrLysAlaLeuIleSerPhe 121
QY 349 ---GAAGGTCAATAT---CTCAAAACATTTGGAAATGTTCTGACCTGAAATTAACCA 402
Db 122 AspaAsnGlyAspTyrPhePro-----ValTyrGlyThrCys---LeuGlyLeuGlu 138
QY 403 TTGAGTGTCTTATTCAGATGACGATTAATTAATTAATGACATATATTTTCTGTACAGAA 462
Db 139 LeuSerValLeuValSerAsnAspAsnLeuThrLeuThrAsnThrSerSerValLys 158
QY 463 ATTCACCTGGAGAAATCAACATCAATCAATGCGGCAAGATTTTAAATTAATGATTC 522
Db 159 LeuProLeuAsnPheThrArgAspSerLysGlnSerArgMetPheArgAsnLeu----- 176
QY 523 TTATTTACCTTATCTGCTGTAACCTGACCTGGAGAGAGATTTTTCAAAACTTTTAC 582
Db 177 -----ProGlugluLeuLeuAsnSerLeuAlaSerGluAsnLeuThrAlaAsnPheHis 194
QY 583 AATGATTAGAGCTAAATCATTAGAAATTATATTGAGAGAAAAAACTTTCTAACT 642
Db 195 LysTP-----SerLeuSerValLysAsnPheThrGluAsnGluLysLysLys--- 211
QY 643 TTCTTTCAGCACCGCAGAGATTAACCTGATGGCAGATAGATTATTTGGCTGGA----- 696
Db 212 PhePheAsnIleLeuThrValAsnThrAspGlyLysThrGluPheIleSerSerMetGlu 231
QY 697 -----CCACAGAGCGCTTAATGAGAGTAGATTTAAGAACTTAATAAAT 744
Db 232 GlyTyrLysTyrProIleTyrAlaValGlnTyrPheProGluLysAlaProPheGluTyr 251
QY 745 AACCAATCAGGAATGATTTTCTATATGAAAGGGGCTGCAAAACAAAGTAT----- 798
Db 252 LysLysLeuArg---GlyIleSerHisAlaProAsnIleValLysThrSerPheTyrLeu 270
QY 799 ---AGTTCATTATTAAGAGGTACAAAGGGTAAAC 831
Db 271 AlaLysPhePheIleSerGluAlaLeuLysAsnAsp 282

RESULT 11
US-09-538-092-119
; Sequence 119, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Glot, Loic
; APPLICANT: Mansfield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CuraPatSeqFormatter Version 0.9
; SEQ ID NO 119
; LENGTH: 623
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
; FEATURE:

```


NAME/KEY: misc feature
LOCATION: (0) (0)
OTHER INFORMATION: Polypeptide Accession Number YDL203C
US-09-538-092-119

Alignment Scores:
Pred. No.: 0.167 Length: 623
Score: 96.00 Matches: 74
Percent Similarity: 36.93% Conservative: 56
Best Local Similarity: 21.02% Mismatches: 114
Query Match: 5.53% Indels: 108
Gaps: 16

US-10-617-962-2 (1-1008) x US-09-538-092-119 (1-623)

```

OY 79 ATACTAACTTTAAGCAACAGATGAG-----GCTCATACAGCATCATAT 123
Db 108 ILePheAsnSerLyProThrGlyLLeGlyTyrAlaAsnAsnGlyThrAsnSerThr 127
OY 124 GGAATTGAATATCGA-----GCTAAGAAATAATATTACTTAC 162
Db 128 GYSerArgTyrGlnLeuProPheAsnPheSerSerThrLySGluSerLeuGlySerPro 147
OY 163 GCTTGCGCTGTAAGTATTCATATGATCTTAACCTTCGATGACTATATTATAGAT 222
Db 148 AlaValGlnAspAlaSerIleSerSerGlyAsnArgIleSerGluSerValArgAspAsn 167
OY 223 -----AAGAGACTGCTGAGAGAAATTATCAAGATAT----- 255
Db 168 SerAlaProProTyrGlnGluSerGluSerArgIleLeuGlnGlnLysValTyrArg 187
OY 256 -----ATGCTCATAT----- 264
Db 188 ThrGlnGlnLysAlaProIleArgProLeuAsnAsnAsnProValProProGlnLysIle 207
OY 265 -----CTTTCATCTGCACTATTAGCGTGAATAATGCGATCAAAATTTCTAAAGT 312
Db 208 AsnGlnProProThrGlySerAlaLysThrAspAsnAsnGlySerSerGlyGlyGlnAsp 227
OY 313 ATGCGAAATGGTTTATTAAGATGAATGATTTTGAAGTCATATCTCAAAACATT 372
Db 228 LysLeuSerSerLySerProGlnLysAlaLeuAlaPheTyrGlnValTyrLysLeuThrIle 247
OY 373 TGGAT-----GTTCTGAGCTTGAAAT-----AAACCATTTGAGTGT 411
Db 248 ThrAspSerSerLyPheThrProGlnLysIleGlnMetGlnTyrCysGlnThrLeuThr 267
OY 412 TATTCAGATGACGATTAATTATTAGCACTATATTTTCTCTGACAGGAAT----- 465
Db 268 TyrAlaPheAsnGlnAspPheIleSerGlnTyrAsnIleAsnAlaGlnLysLeuLysArg 287
OY 466 -----CCACTGAG-----GAAATCAACAATCAAAATCGCGCAAGATTTTAAA 510
Db 288 SerLeuLySerProGlnGlnMetLeuLysAsnGlnLysValIleLeuGlnLysSerPhe 307
OY 511 TTAATGATTTTATTATTAACCTTATCT-----GCTGTAACCTCACTGGAGG--- 558
Db 308 ValLeuThrLysLeuIleThrLeuLysTyrProProAlaMetLysLeuMetGlyThrLeu 327
OY 559 -----AGGATTTTTCAAAAAACTTTTAAACATGATGATAGAGGT 597
Db 328 TyrSerHisGlnProTyrLeuProIleLysAsnLysAsnIleValIleLysAsnAspGln 347
OY 598 AAATCATTAAGAAATTAATTATGAGAAAAAAACTTTCTTAAACCTTTCTTGACACCG 657
Db 348 LysAlaLeuGlnLysTyrCysLysAlaIleLysLeuAsn----- 360
OY 658 CAGAGATTAACCTGATGCGAGATAGTTATTTGGCTGCA----- 696
Db 361 -----AsnSerAspAlaCysTyrArgAlaGlyValCysPheGlnTyrGlnArg 376
OY 697 -----CCAAAGAGCGCCTTAATGAGAGATGAGTTTAAA 732

```

```

Db 377 GlyThrSerSerLeuAspProSerProThrLySGlnGlnCysIleLysLysAlaPheGln 396
OY 733 -----GAACTTAAATAATCAAAATCT-----AGGAATGATTT 765
Db 397 TyrTyrGlnHisGlyAlaGlnValCysSerAsnSerAlaCysMetTyrLysLeuGlyMet 416
OY 766 TCTAATATGAAAGGCGCTGCAAAACAAAG-----TATAGTTCAATT 807
Db 417 SerHisLeuTyrGlyLysLeuAsnMetGlnLysThrAspValLeuLeuAlaIleLysTrpPhe 436
OY 808 ATAAAGAGGTACAAAGGTTACCGCTCCACAGACA 843
Db 437 AspLysAlaAlaGlnLysGlyAspSerProGlnThr 448

RESULT 12
US-08-951-871-2
; Sequence 2, Application US/08951871
; Patent No. 5866398
; GENERAL INFORMATION:
; APPLICANT: XIAO, SHUANG-YONG
; TITLE OF INVENTION: METHOD FOR CLONING
; TITLE OF INVENTION: AND PRODUCING THE BslI RESTRICTION ENDONUCLEASE IN E.
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESS: New England Biolabs, Inc.
; STREET: 32 Tozer Road
; CITY: Beverly
; STATE: MA
; COUNTRY: US
; ZIP: 01915
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/951,871
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Williams, Gregory D
; REGISTRATION NUMBER: 30901
; REFERENCE/DOCKET NUMBER: NEB-131
; TELEPHONE: 978-927-5054
; TELEFAX: 978-927-1705
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 912 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; MOLECULE TYPE: protein
; TOPOLOGY: linear
; FRAGMENT TYPE: internal
US-08-951-871-2

Alignment Scores:
Pred. No.: 0.192 Length: 912
Score: 96.00 Matches: 66
Percent Similarity: 34.84% Conservative: 42
Best Local Similarity: 21.29% Mismatches: 110
Query Match: 5.53% Indels: 92
Gaps: 12

US-10-617-962-2 (1-1008) x US-08-951-871-2 (1-912)
OY 4 GTTATACAAATTAAACCTGATGATGAAGTGATATCCACCGCTTGAAAGCAATATAGCA 63

```

```
Db      21 ValValSerLeuIlaAspGluAsp-----AlaLeuSerLeuGlu 33
Oy      64 GGAGATATATAGTACGTATACAACTTAAAGCAAAACAGATGAGGTCATACAGCATCATAT 123
Db      34 GluGlnPheProLeuAlaLeuSerHisProValValAspGlu----- 47
Oy      124 GGAATGTAATATGAGCTTAAAGAAATATATATAGCTTACGCTTTGGCTGAAGTGATAT 183
Db      48 -----GluIleValIleTyrSileLeuCysGluTyrSileAsnLeuAsnAlaLeuAsnVal 65
Oy      164 CATATGATCTTAAACCTTCATGATGACTTATTAAG-----AATAAAGACTGCT 234
Db      66 LysThrIleSerGluThrLeuAsnLysGluTyrLysPheGluYArgAsnSerLysThrAla 85
Oy      235 GAGAAATTTAT-----CAAGAATATATGCTTAATCTTCATCTGCATCA 279
Db      86 LeuLysLysTyrLeuAspTyrGluLysGluLysIleLeuIleGlnPhePheAsnThrLeu 105
Oy      280 TTAGGTGAAATGTGATCAAAATTTCTAAAGATATGGCAAAATGCTTTTATTAAGATGA 339
Db      106 MetLeuGluAsnAsnThrTyrIleAspArgGlu-----TyrIleGluSerVal 121
Oy      340 CTGGAATTTGAAGTCATATCTCAAAACATTTGGAAATGCTTCAGACTTGAATAA 399
Db      122 LeuAlaPhe-----CysGlu 126
Oy      400 CCATTGAGCTTATTCAGATGACGATAAATTATAGACATATATTTTCTCTGACAG 459
Db      127 ProValSerLysGluLysIleLysAsnGluPheIleLysLeuTyr----- 141
Oy      460 GAAATTTCCAGTGAAGGAAATCAACATCAAAATGCCGAAAGATTTTAAATTAATGAT 519
Db      142 -----AsnGluAlaAsnGluValAsnGluTyrGluLysLeuLysAsp 155
Oy      520 TTCTTATTTACTTATCTGCTGTACTTCACTGGGAGAGATTTTTCAAAACCTTCT 579
Db      156 TyrLeu-----LeuGly-----IleTyrSerLysLeuPhe 165
Oy      580 TACAATGATTAAGAG---GCTAATCATTAAGAAATTAATTAAGAGAAATAAATCTTCT 636
Db      166 SerMetGlyLeuGluAsnLeuArgLeuIleGluIleTyrAsnSerAsnGluSerLeuIle 185
Oy      637 AAACCTTTCTTCGACCCAGCAGAGATTACTGATGGCAGAAATAGTTATTTGGCTGA 696
Db      186 LysLysValPhe----- 189
Oy      697 CCAACGAAAGCCCTAAATGAGAGCTTTTAAAGAACTTAAATAACAATCTAG 756
Db      190 -----LysTyrGluSerThrIleLysGluLeuLysGluTyrCysLeuSer 204
Oy      757 AATGATTTTCTAATATGAGAGGGGCTGCAAAACAAAGATATAGTATTTAATAAGAG 816
Db      205 AsnGlnGluSerIleThrAlaGlyLeuAlaIleLysMetPheAsnGluLysTyrMetGlu 224
Oy      817 GTACAAAAGGATGAGCTTCACAGACAGCAGCGAAAAGTATTTGTCACCCAGTGCCAGT 876
Db      225 LeuMetLysLysGluTyrGlnGlnAspAlaIle-----AlaLeu 237
Oy      877 AACCTGAAAAAATGCCGAATATTTATAT 906
Db      238 LysLeuGlnGluHisMetAsnGlnLeuTyr 247

RESULT 13
US-09-489-039A-14067
; Sequence 14067, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489, 039A
```

```
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 14067
; LENGTH: 3168
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-14067

Alignment Scores:
Pred. No.: 0.301 Length: 3168
Score: 96.00 Matches: 74
Percent Similarity: 33.42% Conservatve: 49
Best Local Similarity: 20.11% Mismatches: 123
Query Match: 5.53% Indels: 122
Db: 4 Gaps: 18

US-10-617-962-2 (1-1008) x US-09-489-039A-14067 (1-3168)
Oy      76 GGTATATCTTAAC-----TTAAGCAAAACAGATGAGGTCATACA 114
Db      981 ArgLeuLeuAsnAsnCysValValAspGlyAspTyrArgCysThrAspGlyArgTyrVal 1000
Oy      115 GCATCATATGGAATTTGAATATCGAGCTTAAGAAAATTAATTTA-----GCTTAC 162
Db      1001 ArgAlaArgProIleGlnHisGlnIleAsnArgGluSerLeuThrGluLeuAlaGlyTyr 1020
Oy      163 GCTTTGGCTGTAAGTGATTT---CATATGATCTTAACCTCGATGACATATATAG 219
Db      1021 CysGluGlyPheGlnAlaIleProAspThrIleAlaArgAlaGlyAspArgLeuTyrGlu 1040
Oy      220 AATAAGAGACTGCTGAGAGA----- 240
Db      1041 MetMetSerGlyAlaGluGluProValAlaIleIlePheProGlnSerAlaSerAspGly 1060
Oy      241 -----ATTATCAAGATATATGTCTAATCTTCACTGCACTTATMGCTGAAT 291
Db      1061 ValGluValLeuTyrGlnGluPheSer-----PheGlyArgTyr 1073
Oy      292 GGTGATTCAAATTTTCMAAGATATGCAAAATGTTTATTAAGATGAATGATTTTGA 351
Db      1074 PheAsnGlnIleAlaIleGlyValLeuArgGlyIleValGlnThrArg-----Gln 1090
Oy      352 GGTCAATATCCTCAAAACATT-----TGG 375
Db      1091 ProArgGlnProLeuArgIleLeuGluValGlyGlyIleThrGlyIleThrAlaTyr 1110
Oy      376 AATGTTCTGAGCTTGAATAATTAACCATTAAGTCTTATTCAGATGACGATTAATTA 435
Db      1111 LeuLeuProGluLeuAsnGlyValProAlaLeuGluTyrHisPheThrAsp---IleSer 1129
Oy      436 GCATCATAT-----TTTTCTCTGTACAG 459
Db      1130 AlaLeuPheThrArgArgAlaGlnGlnLysPheAlaAspTyrAspPheValLysTyrSer 1149
Oy      460 GAAATTCACGTGAGAGAAATCAACATCAATCAATGCC-----GCAAGATTTTAAATTA 513
Db      1150 GluLeuAspLeuGluLysGluAlaGlnSerGlnGlyPheGlnAlaGlnSerTyrAspLeu 1169
Oy      514 ATGATTTTCTTATTAACCTTATCTGCTGTAACTTCATCGGAAAGAGATTTTTCAAA 573
Db      1170 IleValAlaIleAsnValIleHisIleThrArgHisIleGlyArg----- 1184
Oy      574 AACTTTACATATGATTTAGAGCTTAATCATTAAGATTTATATGAGAGAAAACCTT 633
Db      1185 -----ThrLeuAspAsnLeu----- 1189
Oy      634 TCTAACCCTTTCTTCGACCAACCGAGAGATTTA-----CCTGAT 672
Db      1190 ---ArgProLeuLeuLysProGlyLysArgLeuLeuMetArgGluIleThrGlnProMet 1208
Oy      673 GGCAGATATAGTTATTTGGCTGAGACCAACAGAAAGCCCTAATAATGAGAGTGATTTTAA 732
```

Db 1209 ArgLeuPheAspPheValPheGlyProLeuValLeuPro-----1221
QY 733 GAACCTTAATAAATACAAATCTTAGAATGGA-----TTTCTTAATATGGAAGGGCT 783
Db 1222 ---LeuGlnAspLeuAspAlaArgGlnGlyLeuPheLeuThrThrAlaGlnTrpGln 1240
QY 784 GCAAAACAAAGATAGTTCATTCTTAATAAGAGTACAAAAGGCTAGCGCTCACAGACA 843
Db 1241 GlnGlnCysArgGlnAlaGlyPheSerIysValAlaTrp-----LeuProGlnAsp 1257
QY 844 GCAGCAAAAGATATTGTTAGACGCGACAGTGCAGTAACCTGCAAAAATTGCCAATTAATTA 903
Db 1258 GlySerProThrAlaGlyMetSerGlnHisIleIleLeuAlaThrLeuProGlyGlnAla 1277
QY 904 TATAGTGTAGAGCTAGCCAAAACACAGGGTAACTTACT-----945
Db 1278 ValSerAla-----ValThrPheThrAlaProSerGlnProVal 1290
QY 946 -----CAAAATGATACCTGACAT 963
Db 1291 LeuGlyGlnAlaLeuThrAspAsn 1298
RESULT 14
US-09-583-110-3191
; Sequence 3191, Application US/09583110
; Patent No. 6699703
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al.
; TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus
; FILE REFERENCE: Pneumoniae for Diagnostics and Therapeutics
; CURRENT APPLICATION NUMBER: US/09/583,110
; CURRENT FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/107,433
; PRIOR FILING DATE: 1998-06-30
; PRIOR APPLICATION NUMBER: US 60/085,131
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: US 60/051,553
; NUMBER OF SEQ ID NOS: 5322
; SEQ ID NO 3191
; LENGTH: 680
; TYPE: prt
; ORGANISM: Streptococcus pneumoniae
; US-09-583-110-3191
Alignment Scores:
Pred. No.: 0.195 Length: 680
Score: 95.50 Matches: 81
Percent Similarity: 39.89% Conservative: 65
Best Local Similarity: 22.13% Mismatches: 123
Query Match: 5.50% Indels: 97
DB: 4 Gaps: 22
US-10-617-962-2 (1-1008) x US-09-583-110-3191 (1-680)
QY 25 GATAGAGTGAATATCCACCGTTGAAAAGCAATAGACAGATATAGTACTACTA 84
Db 65 AapAlaSerGlyLeuPheLeuValGlnAenThrLeuLys-----GlnValVal 80
QY 85 AACTTAAAGCAACAGATAGGCTCTACAGCATCTATGATGAAATGAAATAGAGCTAG 144
Db 81 SerPheThrArgSerAsnLys-----MetThrAlaThrAspLeuLysGlnThrAlaLys 98
QY 145 AAAATATATAGCTTACGCTTGGCTGTAAGTGAATTCATTAATGATTAAGCTTCT 204
Db 99 Lys---LeuLeuThrTyrlValSerIleSerSerProAsnLeuThrGlnArgGlnLeuAla 117
QY 205 GATGACTATTAATGAATAAGAGACTGCTGAGAGAAATTAATCAAGAAATATATCTAAT 264
Db 118 AspTyrlTyrlLeuAlaAspProGln-----IleTyrlLysThrValGlnAla 133

QY 265 CTT---TCATCTGCATATTAGTGAAAATGGTATCAAAATTTCTTAAGATATGGCAAT 321
Db 134 LeuProSerGlnLysArgLeuAspSerAspIlysAsnArgLeuSerGln-----Ser 150
QY 322 GCTTTTATTAAGAAATGAACCTGATTTTGAAGCTCAATTCCTCAAAACATTTGGAATGT 381
Db 151 GluLeuTyrlAsnAsnAlaValAsp-----158
QY 382 CCGAGCTTGAATAATTAACCATGAGCTTATTCAGATGACGAATTAATTAATGACCTA 441
Db 159 ---SerValGlnThrSerGlnLeuAsn---TyrlThrGlnAspGlnLys---LysGlnIle 175
QY 442 TATTTTCTCT-----GTACAGAAAT 465
Db 176 TyrlLeuPheSerGlnLeuAsnAlaValGlyAsnPheAlaThrGlyThrIleAlaThrAsp 195
QY 466 CCACTGAGCAAAATCAACATCAATCCGACAGATTTTAAATTAATGATTTCTTA 525
Db 196 ProLeuAsnAspSerGlnValAlaValIleAlaSerIleSerLysGlnMetPro-----213
QY 526 TTTACCTATCTGCTGTAACCTCACTGAGAGAGAGATTTTCAAAAACCTTTACAT 585
Db 214 ---GlyIleSerIleSerThrSerTrpAspArgLysValLeuGlnThrSerLeuSer 232
QY 586 -----GATTAAGGCTTAATCATTAAGATTAAT 615
Db 233 IleValGlySerValSerSerGlnLysAlaGlyLeuProIleGlnAlaGlnLysArg 252
QY 616 ATTGAGAGAAAAAACTTTCTTAACCTTTCTTTCGACACCGCAGAGATTACCTGATGC 675
Db 253 LeuLysLys-----GlyTyrlSerLeuAsnAsp---261
QY 676 AGAATAGT-----TATTGGCTGAGACA---ACAGAAGGCCCTAAATGAGAGT 726
Db 262 ArgValGlyThrSerTyrlLeuGlnLysGlnTyrlGlnGlnThrLeuGlnGlyLysArgSer 281
QY 727 TTTAAAGAACTTAATAATTAACCAATCTAGAAATGATTTCTTAATATGAAAGGCTGCA 786
Db 282 ValLysGlnIleLeuAspLys-----TyrlGlyAsnMetGlnSerVal---296
QY 787 AAACAAAAGTATAGTTCATTATTAATAAGAGTACAAAAGGTAAC-----831
Db 297 -----AspThrIleGlnGlnLysSerLysGlyAsnAsnIleLysLeuThr 311
QY 832 -----GCTCCACAGACAGACGCAAAAGT---ATTGTAACGACGACGCTGAC 879
Db 312 IleAspLeuAlaPheGlnAspSerValAspAlaLeuLysSerTyrlPheAsnSerGln 331
QY 880 CTGAA-----AAATTGCCGATATATTTATATAGTGTAGAGCTAGCCAAAAA 927
Db 332 LeuGlnAsnGlyGlyAlaLysTyrlSerGlnGlyValTyrlAlaValAlaLeuAsnProLys 351
QY 928 GACAGGTAACCTTACTCAAAATGATACGACAT-----ACAAATGACG 972
Db 352 ThrGlyAlaValLeuSerMetSerGlyLysHisAspLeuLysThrGlyGlnLeuThr 371
QY 973 GTTCACTAGTGTGGAAT 990
Db 372 ProAspSerLeuGlyThr 377
RESULT 15
US-08-328-322-5
; Sequence 5, Application US/08328322
; Patent No. 5723436
; GENERAL INFORMATION:
; APPLICANT: Huang, Laiqiang
; APPLICANT: Cyert, Martha S.
; TITLE OF INVENTION: Calcineurin Interacting Protein Compositions
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Denlinger & Associates
; STREET: 350 Cambridge Avenue, Suite 250

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: November 21, 2004, 08:44:30 ; Search time 89 Seconds

(without alignments)
8021.587 Million cell updates/sec

Title: US-10-617-962-2

Perfect score: 1735

Sequence: 1 atggttataacattacacc.....ctcattataaaatatataga 1008

Scoring table:
BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 1570615 seqs, 354127592 residues

Total number of hits satisfying chosen parameters: 3141230

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Command line parameters:
-MODEL=frame+ n2p.model -DEV=xlp
-O=/cgn2_1/USPTO.spool.P/US10617962/rnat.17112004.163923.9785/app.query.fasta_1.1159
-DB=Published Applications AA -QMT=faeascan -SUFFIX=xabp -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MOD=LOCAL -OUTFMT=pro -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USERS=US10617962.@cgn_1.1.130.@rnat.17112004.163923.9785
-NCPU=6 -ICPU=3 -NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications AA:*

1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep:*
17: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
18: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep:*
19: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep:*
20: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result

Query

No.	Score	Match	Length	DB	ID	Description
1	1734	99.9	335	15	US-10-617-962-4	Sequence 4, Appli
2	1276	73.5	368	15	US-10-617-962-3	Sequence 3, Appli
3	110	6.3	652	17	US-10-437-963-149602	Sequence 149602,
4	102.5	5.9	1581	17	US-10-425-115-293701	Sequence 293701,
5	101.5	5.9	1639	14	US-10-087-164-10	Sequence 10, Appli
6	100.5	5.8	1921	17	US-10-425-115-293718	Sequence 293718,
7	100	5.8	297	17	US-10-425-115-271828	Sequence 344697,
8	98.5	5.7	367	14	US-10-369-493-197667	Sequence 197667,
9	98.5	5.7	616	15	US-10-282-122A-76864	Sequence 76864, A
10	98.5	5.7	1162	16	US-10-437-963-200517	Sequence 200517,
11	98	5.6	580	15	US-10-282-122A-54429	Sequence 54429, A
12	98	5.6	954	15	US-10-282-122A-63608	Sequence 63608, A
13	97.5	5.6	1770	17	US-10-425-115-271828	Sequence 271828,
14	97	5.6	878	17	US-10-425-115-249308	Sequence 249308,
15	97	5.6	1477	16	US-10-437-963-115886	Sequence 115886,
16	96.5	5.6	685	9	US-09-815-242-13477	Sequence 13477, A
17	96.5	5.6	685	15	US-10-282-122A-74418	Sequence 74418, A
18	96.5	5.6	685	15	US-10-282-122A-54288	Sequence 54288, A
19	96	5.5	879	15	US-10-282-122A-54288	Sequence 192563,
20	96	5.5	1006	16	US-10-437-963-192563	Sequence 192563,
21	95.5	5.5	303	15	US-10-282-122A-51836	Sequence 51836, A
22	95.5	5.5	360	15	US-10-282-122A-54381	Sequence 54381, A
23	95.5	5.5	414	14	US-10-369-493-22593	Sequence 22593, A
24	95.5	5.5	653	16	US-10-437-963-186821	Sequence 186821,
25	94.5	5.4	507	15	US-10-425-114-57458	Sequence 57458, A
26	94.5	5.4	611	15	US-10-425-114-54378	Sequence 54378, A
27	94.5	5.4	650	17	US-10-425-115-353793	Sequence 353793,
28	94.5	5.4	696	9	US-09-817-647-23	Sequence 23, Appli
29	94.5	5.4	696	9	US-09-877-665-23	Sequence 23, Appli
30	94.5	5.4	696	13	US-10-136-573A-23	Sequence 23, Appli
31	94.5	5.4	696	14	US-10-215-862-23	Sequence 23, Appli
32	94.5	5.4	1692	17	US-10-425-115-21469	Sequence 21469,
33	94	5.4	386	15	US-10-282-122A-54544	Sequence 54544, A
34	94	5.4	534	9	US-09-801-368-124	Sequence 124, App
35	94	5.4	737	16	US-10-437-963-107270	Sequence 107270,
36	94	5.4	910	15	US-10-456-129-7	Sequence 7, Appli
37	94	5.4	1009	15	US-10-607-631-2	Sequence 2, Appli
38	94	5.4	1097	14	US-10-369-493-1577	Sequence 1577, Ap
39	93.5	5.4	344	15	US-09-820-843A-35	Sequence 35, Appli
40	93.5	5.4	411	15	US-10-282-122A-53440	Sequence 53440, A
41	93	5.4	924	14	US-10-369-493-18383	Sequence 18383, A
42	93	5.4	1830	15	US-10-424-599-147146	Sequence 147146,
43	93	5.4	2872	17	US-10-783-528-116	Sequence 116, App
44	93	5.4	6713	15	US-10-282-122A-43811	Sequence 43811, A
45	92.5	5.3	539	16	US-10-408-765A-1437	Sequence 1437, Ap

ALIGNMENTS

RESULT 1
US-10-617-962-4
; Sequence 4, Application US/10617962
; Publication No. US2004005036A1
GENERAL INFORMATION:
APPLICANT: Commonwealth Scientific and Industrial Research Organisation
APPLICANT: EAST, Peter David
TITLE OF INVENTION: Toxin Genes from the Bacteria Xenorhabdus nematophilus and Photo:
FILE REFERENCE: 050179-0076
CURRENT APPLICATION NUMBER: US/10/617,962
CURRENT FILING DATE: 2003-07-14
PRIOR APPLICATION NUMBER: US/09/463,048A
PRIOR FILING DATE: 2002-12-13
PRIOR APPLICATION NUMBER: PCT/AU98/00562
PRIOR FILING DATE: 1998-07-17
PRIOR APPLICATION NUMBER: PO 8088
PRIOR FILING DATE: 1997-07-17
NUMBER OF SEQ ID NOS: 20
SOFTWARE: PatentIn version 3.1
SEQ ID NO 4
LENGTH: 335

```
; TYPE: PRT
; ORGANISM: Photorhabdus luminescens
US-10-617-962-4

Alignment Scores:
  Pred. No.:      3,116-161      length:      335
  Score:          1734.00        Matches:    335
  Percent Similarity: 100.00%    Conservative: 0
  Best Local Similarity: 100.00%  Mismatches: 0
  Query Match:     99.94%       Indels:     0
  DB:              15           Gaps:         0

US-10-617-962-2 (1-1008) x US-10-617-962-4 (1-335)

QY      1 ATGGTATTAACAATTAAACCTGATGATAGAGTATATCAACCCGTTGAAAAACAATA 60
DB      1 MetValIleGlnLeuThrProaspAspArgSerGlyTyrProProValGlnIuysGlnIle 20
QY      61 GCAAGAGATATGTCAGTACTTAAACCTTAAAGCAACAGATGAGGTCATACAGCATCA 120
DB      21 AlaGlyAspIleValArgIleLeuAsnPhelySGlnThrAspGlnGlyIleThrAlaSer 40
QY      121 TATGGAATTGAATATCGAGCTAAGAAAAATATATTAGCTTAGCTTTGGCTGTAAGTGT 180
DB      41 TyrGlyIleGlnTyrArgAlaIuysIleIleLeuAlaTyrAlaLeuAlaValSerGly 60
QY      181 ATTCATAATGTATCTAAACCTTCCTGATGACTATTATAGAAATAAGAGACTGTCAGAGA 240
DB      61 ILeHAsnValSerIleuLeuProaspAspTyrTyrIuysAsnIuysGlnThrAlaGlnArg 80
QY      241 ATTATCAAGATATATGCTTAATCTTTTCATCTGCACCTTTAGGTAAGAAATGGTGATCA 300
DB      81 ILeTyrGlnIuysTyrMetSerAsnLeuSerSerAlaLeuLeuGlyGlnIuysngIuysGln 100
QY      301 ATTTCTAAAGATATGCGCAATGTTTCTTTTAAAGAAATGAACGTGATTTGAAGTCAATAT 360
DB      101 ILeSerIuysAspMetAlaAsnngIuysTyrIuysAsnGlnIuysAspPhelyGlnIuysTyr 120
QY      361 CCTCAAAACATTGGAAATGTCCTGAGCTTGAATAATTAACATGAGTCTTATTCAGAT 420
DB      121 ProGlnAsnIleTyrAsnValProGlnIuysGlnIuysLeuProIuysSerAlaTyrSerAsp 140
QY      421 GACGATAAATTATTAGCACTATATTTTCTCTGTCAGAGAAATTCACCTGAGAGAAAT 480
DB      141 AspAspIuysIleuLeuAlaLeuTyrIuysPheSerValGlnGlnIleProIuysGlnIuysAsn 160
QY      481 CAACAATCAAAATGCGCAAGATTTTAAATTAATTGATTTTCTTATTAACCTTAATCTGT 540
DB      161 GlnGlnIuysAsnAlaAlaIuysPhePheIuysLeuIleAspPheIuysPheThrIuysSerAla 180
QY      541 GTAACCTCACTGCGAAGAGATTTTCTTCAAAAACCTTTACAGATGATTAAGGCTAAA 600
DB      181 ValThrSerLeuGlyIuysArgIlePheSerIuysAsnPhelyIuysGlnIuysAlaIuys 200
QY      601 TCATTAGAGAAATTATATGAGAGAAAAAACTTTCTTAAACCTTTCTTGCACCAACGCGAG 660
DB      201 SerIuysGlnIuysTyrIleGlnIuysIuysLeuSerIuysProPhePheArgProProGln 220
QY      661 AGATTACCTGATGCGAAGATAGTTATTTGGCTGCGACCAAGAGCGGCTTAATGAGAA 720
DB      221 ArgIuysProaspGlyIuysArgIleGlyTyrIleuAlaGlyProThrGlnIuysAlaProIuysTyrParg 240
QY      721 GTGAGTTTAAAGAACTTAAATAACAATCTAGAGATGATTTTCTTATATGAGAGGG 780
DB      241 ValSerPheIuysGlnIuysAsnIuysAsnIuysSerArgAsnGlyPheSerAsnIuysGlnIuys 260
QY      781 GCTGCAAAACAAGATATGCTTATTAATAAGAGCTACAAAGGTAACGCTCCACAG 840
DB      261 AlaAlaIuysGlnIuysTyrSerSerPheIleIuysGlnIuysAlaIuysGlnIuysAlaProGln 280
QY      841 ACAGAGAGGAAAGATTTGCTACAGCAGTGCAGTAACCTGAGAAATTTGCGCAATAT 900
DB      281 ThrAlaAlaIuysSerIleGlyThrAlaSerGlySerAsnIuysGlnIuysLeuProAsnAsn 300
```

```
QY      901 TTATATAGTGTAGGCTTAAGCCAAAAAGACAGGGTAACCTTACTCAAAATGATATCTGAC 960
DB      301 LeuTyrSerValArgLeuSerGlnIuysAspArgValThrPheThrGlnAsnAspThrAsp 320
QY      961 AATACAAATGACGCTTCATAGTGTTCGACATCTTATTAATAATATA 1005
DB      321 AsnThrMetThrValIleSerValGlnIuysThrIuysTyrIuysAsnIle 335

RESULT 2
US-10-617-962-3
; Sequence 3, Application US/10617962
; Publication No. US2004005036X1
; GENERAL INFORMATION:
; APPLICANT: Commonwealth Scientific and Industrial Research Organisation
; APPLICANT: EAST, Peter David
; TITLE OF INVENTION: Toxin Genes from the Bacteria Xenorhabdus nematophilus and Phocor
; TITLE OF INVENTION: luminescens
; FILE REFERENCE: 050179-0076
; CURRENT APPLICATION NUMBER: US/10/617,962
; PRIOR FILING DATE: 2003-07-14
; PRIOR APPLICATION NUMBER: US/09/463,048A
; PRIOR FILING DATE: 2002-12-13
; PRIOR APPLICATION NUMBER: PCT/AU98/00562
; PRIOR FILING DATE: 1998-07-17
; PRIOR APPLICATION NUMBER: PO 8088
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 368
; TYPE: PRT
; ORGANISM: Xenorhabdus nematophilus
US-10-617-962-3

Alignment Scores:
  Pred. No.:      2,99-116      length:      368
  Score:          1276.00        Matches:    368
  Percent Similarity: 80.17%    Conservative: 36
  Best Local Similarity: 70.11%  Mismatches: 47
  Query Match:     73.54%       Indels:     24
  DB:              15           Gaps:         3

US-10-617-962-2 (1-1008) x US-10-617-962-3 (1-368)

QY      4 GTTATACATTAACAACCTGATGATGA-----AGTGATATCCACC 45
DB      11 ValIleGlnLeuThrProaspAspArgValThrProaspAspIuysGlyIuysTyrGlnPro 30
QY      46 GTTGAAGAAACAATTAAGAGAGATATAGCTATATCAAACTTTAAGCAAAACAGATGAG 105
DB      31 ValGlnIuysGlnIleAlaGlyAspIleIleArgValLeuGlnIuysPheIuysGlnIuysAsnGln 50
QY      106 GGTCAATACAGCATCATATGAAATTAATATGAGCTTAAGAAAAATAATATTAGCTTACGCT 165
DB      51 SerIleThrIuysIleuTyrGlyIleAlaTyrIuysArgAlaIuysValIleIleAlaTyrAla 70
QY      166 TTGGCTGTAAGTGAATTCATATGATATGATCTTAACTTCGATGACTTATTAAGATAA 225
DB      71 LeuAlaValSerGlyIleIleAsnValSerGlnIuysLeuProGlnIuysPheTyrIuysAsnIuys 90
QY      226 GAGACTGCTGAGAGAAATTTATCAAGAAATAATGTCTAATCTTTCTGCACTATTAAGT 285
DB      91 AspAsnThrIuysArgIleTyrGlnIuysTyrMetSerAsnIuysLeuSerAlaLeuIuysGly 110
QY      286 GAAATATGATCAAAATTTCTTAAGATATGCAAAATGGTTTATTATTAAGATAAATCTGAT 345
DB      111 GlnAsnGlyAspGlnIleSerIuysAspMetAlaAsnAspPheThrGlnAsnGlnIuysGln 130
QY      346 TTTGAAGGCAATATCCCAAAACATTTGGAATGTTCTCGAGCTTGAATAATTAACCATTTG 405
DB      131 PheGlyIuysGlnIuysArgLeuIuysAsnThrThrAspIleProaspLeuGlnIuysLeuIuys 150
```



```

QY 406 AGTCTTATGATGAGATTAATTTATGACATAATTTTCTGTACAGAAATT 465
    |||
    |||
    |||
Db 151 GLUspTySerAspGluAspLysLeuLeuLeuLeuLeuLeuLeuLeuLeu 170
    |||
    |||
    |||
QY 466 CCAGTGGAGAAATCAACATCAATGCCGACAGATTTTAAATTAATGATTTCTTA 525
    |||
    |||
    |||
Db 171 PrometGluAlaAsnGlnGlnSerAsnAlaAlaAsnPhenylValIleAspPheLeu 190
    |||
    |||
    |||
QY 526 TTTACCTTATGCTGCTTACTTCACTGGAGAGAGATTTTTCAGAAAACCTTTACAT 585
    |||
    |||
    |||
Db 191 LeuIleuSerAlaValThrSerLeuGlySarGlyIlePheSerLysAsnPhenylVal 210
    |||
    |||
    |||
QY 586 GGATTAGAGCTTAATCATTAAGATTAATTAAGAGAGAGAGAGAGAGAGAGAGAGAG 645
    |||
    |||
    |||
Db 211 GlyLeuGlnThrLysSerLeuGlnLysenryIleGluAlaGlyLysLeuSerLysProhe 230
    |||
    |||
    |||
QY 646 TTTGACCAACCGGAGAGATTAACCTGATGCGAGAGATAGGTTATTTGGCTGACCAAGAA 705
    |||
    |||
    |||
Db 231 PheArgProGlnLysLeuProAspGlyArgThrGlyLysLeuAlaGlyProThrLys 250
    |||
    |||
    |||
QY 706 GCGCTAA----- 714
    |||
    |||
    |||
Db 251 AlaProLysLeuProThrThrSerSerThrAlaThrThrSerThrAlaAlaSerSerAsn 270
    |||
    |||
    |||
QY 715 TGAAGAGTGAAGTTTAAGAACTTAATAATAACAATCTAGAGATGATTTTCTAATAG 774
    |||
    |||
    |||
Db 271 ThrArgValSerLeuGlnLysLeuArgAspAsnProSerArgAsnThrPheMetLysMet 290
    |||
    |||
    |||
QY 775 GAAGGGCTGCAAAACAAAGTATAGTTCATTATTAAGAGTACAAAGGATACCT 834
    |||
    |||
    |||
Db 291 AspSerAlaAlaLysArgLysSerSerPheIleLysGlyValGlnLysGlyLysAsp 310
    |||
    |||
    |||
QY 835 CCAGAGACAGAGCGAAAGATTTGGTACAGCCATGCGATACCTGGAATAATGGCG 894
    |||
    |||
    |||
Db 311 ProAlaAlaAlaAlaSerIleGlyThrLysSerGlySerAspPheGlnLysLeuGln 330
    |||
    |||
    |||
QY 895 ---AATAATTAATAGTGTGAGCTAAGCCAAAGAGAGGTAACCTTTACTCAAAAT 951
    |||
    |||
    |||
Db 331 GlyArgAspLeuLysSerIleArgLeuSerGlnGlnLysArgValThrPheSerLysLeu 350
    |||
    |||
    |||
QY 952 GATATGACAAATACATGACGATGAGTGTGAGTGAATCATTATTAATAATTA 1005
    |||
    |||
    |||
Db 351 AsnThrAspGlnIleMetGlnIleGlnSerValGlyThrHisLysGlnAsnIle 368
    |||
    |||
    |||

```

```

Score: 110.00 Matches: 81
Percent Similarity: 33.25% Conservative: 56
Best Local Similarity: 19.66% Mismatches: 104
Query Match: 6.34% Indels: 171
DB: 16 Gaps: 23
US-10-617-962-2 (1-1008) x US-10-437-963-149602 (1-652)
QY 127 ATTAGAATATGAGCTTAAGAAATATATATATATATACCTTACGCTTTGGCT-----GTA 174
    |||
    |||
    |||
Db 1 MetAspLeuLysThrIleLysIleLeuLeuArgTyrProArgGlySerAspGlyLysLeu 20
    |||
    |||
    |||
QY 175 AGTGTATTCATTAATGATATCTTAACCTTCGATGACATTAATTAAGATTAAGAGAGAGAG 234
    |||
    |||
    |||
Db 21 AlaGlyValHisGlyPheLeuGlnPheAla-----TyrLysAspLysSerGlnAsp 37
    |||
    |||
    |||
QY 235 GAGAGATTTTCAAGAAATATATATATATATCTTATCTTCACTGACCTTTAGGAGAAATGCT 294
    |||
    |||
    |||
Db 38 ThrLysIleTyrCysProCysGlnThrCysValHisThrMetLeuLeuSerLysAsn--- 56
    |||
    |||
    |||
QY 295 GATCAAAATTTCTAAAGATATGCAAAATGTTTATATAG-----AATGACTGATTTT 348
    |||
    |||
    |||
Db 57 AspValTyrAspArgLeuValCysAsnGlyMetLeuGlnInseryTyrAspGlnThrAspPhe 76
    |||
    |||
    |||
QY 349 GAAGGT-----CAATATCTCAAAACATTTGGAATGTT 381
    |||
    |||
    |||
Db 77 HisGlyLysSerSerGlnGlnGlnThrArgAsnGlnGlnProGln----- 91
    |||
    |||
    |||
QY 382 CCGAGCTTGAATAATTAACATGAGTCTTATTCAC---GATGACGATTAATTAATTAAGCA 438
    |||
    |||
    |||
Db 92 ProHisAsnGlnAsnMetArgAlaAsnMetHisGlnLeuIleAspAspAlaAspPheArgThr 411
    |||
    |||
    |||
QY 439 CTATATTTTCTCTGTACAGAAATTCACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 489
    |||
    |||
    |||
Db 112 ValTyr-----AspAspValProMetSerAspHisAlaAspSerProCysThr 519
    |||
    |||
    |||
QY 490 -----AATGCCGAGAGATTTTAAATTAATTAATGAT----- 519
    |||
    |||
    |||
Db 128 HisThrAspGlyProAsnLeuGlnAlaGlnAlaPheTyrLysLeuValLysAspSerGln 147
    |||
    |||
    |||
QY 520 -----TTCTTATTTAC 531
    |||
    |||
    |||
Db 148 LysProLeuTyrAspGlyCysGlnLeuSerHisLeuSerLeuLeuValLeuLeuPheAsn 167
    |||
    |||
    |||
QY 532 TTATCTGCTGTAACCTTCACTGGAAGAGAGATTTTTCAC----- 570
    |||
    |||
    |||
Db 168 MetLysSerMetAsnLysTyrPheSerAspLysSerPheGlyAspLeuLeuAspIleLeuHis 187
    |||
    |||
    |||
QY 571 -----AAAACTTTTACAAAT----- 585
    |||
    |||
    |||
Db 188 MetAlaIleProAsnGlnLysGlnLeuProLysAsnPhenylValAlaLysLysValVal 207
    |||
    |||
    |||
QY 586 -----GATTAGAG----- 594
    |||
    |||
    |||
Db 208 SerLysPheGlyLeuAspTyrGlnGlnIleHisAlaCysProAsnAsnCysGlnLeuPhe 227
    |||
    |||
    |||
QY 595 -----GCT 597
    |||
    |||
    |||
Db 228 TrpLysAspLysValAsnAspAspPheCysSerIleCysLysAlaSerArgTyrLysGly 247
    |||
    |||
    |||
QY 598 AAATCATTAAGAAATTAATATTT-----GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 630
    |||
    |||
    |||
Db 248 LysLysProGlnThrLysLeuThrLysSerLysLysLysValAlaThrProSerLysValLeu 267
    |||
    |||
    |||
QY 631 -----CTTTCT 636
    |||
    |||
    |||
Db 268 ArgTyrPheProIleLysAspArgLysArgLeuPheMetCysArgGlnThrAlaSer 287
    |||
    |||
    |||
QY 637 AAACCTTTCTTTCAGACACCGAGAGATTAACCTGATGAGAGATGATGATTTGGCTGGA 696
    |||
    |||
    |||
Db 288 LeuProAlaGlyTyrPheAspGlnGlnLysLysLysLysLysLysLysLysLysLysLys 304
    |||
    |||
    |||
QY 697 CCAGAGAGAGCGCTTAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 747
    |||
    |||
    |||

```

Alignment Scores: 0.0968 Length: 652

Score: 101.50 Matches: 82
 Percent Similarity: 34.70% Conservative: 53
 Best Local Similarity: 21.08% Mismatches: 137
 Query Match: 5.85% Indels: 117
 DB: 14 Gaps: 19

US-10-617-962-2 (1-1008) x US-10-087-464-10 (1-1639)

```

QY 121 TATGAAATTAATGAGCT-----AAGAAATAATATTAGCTTAC 162
DB      ::::::::::::::|
QY 205 Pheasnleuylsileaylalaasnleuleuaspvalleuleylsleuvalgheglytr 224
DB      ::::::::::::::|
QY 163 GCTTGGCTGTAAGTGTATT---CATATGTATTAACCTCCGATGACTATTATTAAG 219
DB      ::::::::::::::|
QY 225 Arglysrproleuaspheasnilelysaspasnvalglylsmetgluaptryllelysls 244
DB      ::::::::::::::|
QY 220 AATTAAGACTGCTGAG----- 237
DB      ::::::::::::::|
QY 245 Asnlyslslethrllegluasnileasnleuleuilegluuserlyelylethrlleasp 264
DB      ::::::::::::::|
QY 238 -----AAGATTTATCA--GAATATATG 258
DB      ::::::::::::::|
QY 265 Lysasnlysaasnalaathlysgluclugluylslyslsleuylrglualeglntryasp 284
DB      ::::::::::::::|
QY 259 TCTAATCTTCATCTGCACATTAAGTGAATAAGTGTATCAAAATTTCT-----AAGAT 312
DB      ::::::::::::::|
QY 285 Leuethrleryrasnlysglnleugluahlsasnleullesevalleugluys 304
DB      ::::::::::::::|
QY 313 ATGGCAAAATGTTTTTAATGAATGA-----CTGAT-- 345
DB      ::::::::::::::|
QY 305 Arglyleapthrlleuleylsasnleuasnilelysgluuleuaspylelleasnleu 324
DB      ::::::::::::::|
QY 346 TTTGAAGTCAATATCTCAAAACCTTTGGAATGTTCTGAGCTT-----GAATAT 396
DB      ::::::::::::::|
QY 325 Ilelysaenproproalaasnserglyasnthrproasnthrlleuaspyleasn 344
DB      ::::::::::::::|
QY 397 AAACCATGAGTGTATTATTCAGATACGATTAATATTATGACATATTATTT---TTCCT 453
DB      ::::::::::::::|
QY 345 Lyslyslleeglunhlsleuilelysgluilehlyethrllelyepheasn 364
DB      ::::::::::::::|
QY 454 GTACAGGAATT-----CCACTG----- 471
DB      ::::::::::::::|
QY 365 Ileapserleupheathrasprouleugluuyltryrlyeuargluylsasnlys 384
DB      ::::::::::::::|
QY 472 -----GAGAAATCAACAATCA 489
DB      ::::::::::::::|
QY 385 Asnileaspileseralylevalgluthrysgluserthrghuiproasnleuylrpro 404
DB      ::::::::::::::|
QY 490 AATGCCGCAAGATT---TTTAATTAAATGATTCTTATTACCTTAATCTGCTGAAT 546
DB      ::::::::::::::|
QY 405 Asnlyslvalthrtryproleusertryrasnspilleasnvalleuasnleuasn 424
DB      ::::::::::::::|
QY 547 TCACTGGAGAGAGATTTT-----TCAAAAACCTTTTAC 582
DB      ::::::::::::::|
QY 425 Serpheglyaspheulleasnproheasptryrthrysgluuproserlyseasnlelyr 444
DB      ::::::::::::::|
QY 583 AATGATTAAGAGCTTAATCATTAAGAAATTAATGAGAAA---AACTTTCTTAA 639
DB      ::::::::::::::|
QY 445 Thrspasnlgularglyleypheulleasnleuileysgluylsillelysllegluys 464
DB      ::::::::::::::|
QY 640 CCTTCTTCACACACCCAGAGATTACCTGATGCAATATGATTTATTTGGCTGACCA 699
DB      ::::::::::::::|
QY 465 Lyslyslleegluseraspyleyseryrghuapshgserlyserleuasnaspille 484
DB      ::::::::::::::|
QY 700 ACAGAAGCGCTAAATGAGAGTGAATTAAAGACTTAATAATACAAATCTAGAAAT 759
DB      ::::::::::::::|
QY 485 Thrlysls-----gluylrgluylsleuasnleuasnleuyltryraserlyseasn 502
DB      ::::::::::::::|
QY 760 GGA-----TTTTCAATATGGAAGGCTGCAAAACAAAGTAAATGTTATTAATA 813
DB      ::::::::::::::|
QY 503 Asnileapleuthrasnphleuylslyswetwetyllyrargtrysertrylysevalgu 522
DB      ::::::::::::::|
QY 814 GAGGTACAAAAGGTAAAGCTCCACAGACAGCAAGAAAGTATTGTACAGCCAGTGC 873
  
```

```

DB      ::::::::::::::|
DB 523 Lysleuthrhlshlsasnthrphlealasertrygluasn-----Serlysls 537
QY 874 AGTAACTCGAATAAATTCGCCGAATTAATTA-----TATAGCTGAGCTA 918
DB      ::::::::::::::|
DB 538 Hisasnleuugluylsleuthrlyslaleuylslyrhetgluaptryrserleuarg-- 556
QY 919 AGCCAAAAGACAGGTTAACTTACTCAAAATGATCTGACATPACATGACGGTTTAT 978
DB      ::::::::::::::|
DB 557 -----Asnilevalvalglu 561
QY 979 AGTGTGAACCTCATTAATAAATATA 1005
DB      ::::::::::::::|
DB 562 Lysgluleuylslyrtrylyllysaasnleu 570

```

RESULT 6

```

US-10-425-115-293718
; Sequence 293718, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(5322)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 293718
; LENGTH: 1921
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MNT4577_30954C.1.pdp
US-10-425-115-293718

```

Alignment Scores:

Pred. No.:	Length:
1.08	1921
Score: 100.50	Matches: 69
Percent Similarity: 36.98%	Conservative: 46
Best Local Similarity: 22.19%	Mismatches: 105
Query Match: 5.79%	Indels: 91
	Gaps: 16

```

US-10-617-962-2 (1-1008) x US-10-425-115-293718 (1-1921)
QY 103 GAGGTCAATACAGCATATGAAATGAATAT----- 135
DB      ::::::::::::::|
QY 1173 Glnelyltryrghlnleeglulyleuasphegllygluthrphlealaprovalalaarg 1192
DB      ::::::::::::::|
QY 136 ---CGAGCTAAGAAATTAATTAATTAATGCTTGGCTGTAAGTGTATTCAATATGTA 192
DB      ::::::::::::::|
QY 1193 Leuclualatlearglleleuualatylrlylacyala-----Hisasnle 1208
DB      ::::::::::::::|
QY 193 TCTAACTCTCATGCTAT-----TATAAGATTAAGACACTGCAAGAAAT 243
DB      ::::::::::::::|
QY 1209 Lysleuylrglnmetaspvallysevalapheleuasnleuyllylleserleuval 1228
DB      ::::::::::::::|
QY 244 TATCAAGAA-----TATATCTTAATCTT 267
DB      ::::::::::::::|
QY 1229 Tyrvalgluglnproproglyrphlegluasproulysarproasnhlvalrphelysleu 1248
DB      ::::::::::::::|
QY 268 TCAATCTGCATTAAGTGAATAATGTCATCAATTTCTAAAGATTAAGCAATGCTTTT 327
DB      ::::::::::::::|
QY 1249 Serlyslaleuylrgly-----leuylsglnalaproargalatr 1262
DB      ::::::::::::::|
QY 328 TATAAGATTAAGCTGAATTTTGAAGTCAATATCTCAAAACATTTGGAATGTT---CCT 384
DB      ::::::::::::::|
QY 1263 Tyrlylrgleuarlgapheleuuserlysaapheyllellyllysevalaspthr 1282
DB      ::::::::::::::|
QY 385 GAGCTTGAAATTAACCATTTG-----AGTGTATTATTCAGATGAC 423
  
```

```
Db      1263 ThrLeuPheThrLysArgIleGlyLysAspLeuPheValCysGlnIleTyrValAspAsp 1302
QY      424 GATTAATATTATGACCTATATTTTCTCTGACGAAATTCACCTGGAGGAAATCA 483
Db      1303 -----IleIlePheGlySerThrAsnGlnLeuPheCysIleGlu----- 1315
QY      484 CAATCAATGCCGACAGATTTTAAATTAATTGATTTCTTATTTACCTTATCTGCTGA 543
Db      1316 -----PheGlyLysMetMetSerLysGlnPheGlnIleSerMetIle 1329
QY      544 ACTTCACCTGGAGAGGATTTTTCAAAACCTTTACATGAGATTGAGGCTAAATCA 603
Db      1330 GlyGluLeu-----SerPhePheLeuGlyLeuGlnIleLysGln 1342
QY      604 TTAGAGAAAT-----TATATTGAGAGAAAAAATTTCTTAACCTTTCTTGA----- 651
Db      1343 LeuLysAspGlyIlePheIleSerGlnSerLysTyrLeuLysAspMetLeuLysPhe 1362
QY      652 -----CCACCGCAGAGATTACCTGATGCGAGAAATAGTTATTG----- 690
Db      1363 GlyLeuGlnAsnAlaLysProIleLysThrProMetAlaThrAsnGlyHisLeuAspLeu 1382
QY      691 -----GCTGGACCAACAGAAAGCGCCTAAATGAGAGTGAATTAAAGA 735
Db      1383 AspGluGlyValAlaCysGlyGlnSerAspSerLeuGln-----GlnHis 1397
QY      736 CTTAAATATACAAATCTAGAGATTTTCTTAATATGAGAGGCGTGCACAAACAAAG 795
Db      1398 SerLeuAsnValSerSerSerArgLysPheGly-----GlyLeuArgLysLysLeu 1414
QY      796 TATAGTCTTATTAATAAGAGGTACAAAAGGTAAC-----GCTCCACAGACA 843
Db      1415 AlaGlyArgPheLysSerLysArgProArgGlyAsnAspAspArgLysThrProThrThr 1434
QY      844 GCAGCGAAAGATTTGGTACAGCAGCTGCGCAGT 876
Db      1435 AspSerGlnAlaGlnSerSerAlaGlyGlyThr 1445

RESULT 7
US-10-425-115-344697
; Sequence 344697, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 344697
; LENGTH: 297
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_77529C.1.pcp
US-10-425-115-344697

Alignment Scores:
Pred. No.: 0.763 Length: 297
Score: 100.00 Matches: 58
Percent Similarity: 34.16% Conservative: 25
Best Local Similarity: 23.87% Mismatches: 76
Query Match: 5.76% Indels: 84
DB: 17 Gaps: 12

US-10-617-962-2 (1-1008) x US-10-425-115-344697 (1-297)
QY      145 AAAATAATATTAGCTTACGCT-----TTGGCTGTAAAGTGATTCATAATGATATCT 195
```

```
Db      95 LysValThrLeuAlaIleAlaIleAspAspLeuCysLeuIleGlyPheGlnAsnLaser 114
QY      196 AAACCTTCGTAGTACTATATTAAGAAATTAAGAGACTGCTGAGAGATTTATCAAGAAATAT 255
Db      115 -----AspHisTrpTyrLys----- 119
QY      256 ATGTCTAATCTTTCACTGTCGACTATTAGGTAAAAATGCT-----GAT 297
Db      120 PheAsnGlyGlnSerSerPheLysGlyLeuProGlyAlaThrValLeuProIleArg 139
QY      298 CAATTTTAAAGATATGCGCAAGTGTTTTATAAGATGAACGATTTTGAAGCTCA 357
Db      140 GlnAsnTyrGlnAspLeuIleLysGlyHis----- 149
QY      358 TATCTCAAAACATTTGGAAATGTTCTGAGCTGGAATAAATCAATGATGCTTATTC 417
Db      150 -----AlaAsnLeuTrpLysValProLeuGlyLysSerAlaIleHisAla----- 165
QY      418 GATGAGATTAATATTATGCACTATATTTTCTCTGTA---CAGAAATTCACCTGAG 474
Db      166 -----ThrLysGlnLeuAlaThrTyrAspArgAlaValThrProAspSerGlnLeuLys 183
QY      475 GAAAT-----CAACAATCAATGCCGAGATTTTAAATTAAT 516
Db      184 AspGlyLeuAlaArgPheValAlaMetCysGlyGlyMetArgPheArgSerIleArg 203
QY      517 GATTTCTTATTTATTCCTTATCTGCTGTAACTTCACTGGAGAGAGATTTTTCAAAAC 576
Db      204 AspMetPheSerSerLeuSer-----GlyAsnAsnTrpGlnGlnLur 218
QY      577 TTTTCAATGATTTAGAGCTTAATCATTAAGAAATTTATGAGACAAAAAATCTTCT 636
Db      219 PheIleThrGlnLeuGlnAlaLysSerValAlaTyrTrpSerGln----- 233
QY      637 AAACCTTCTTTCAGCAGCAGAGATTAACCTGATGCGAAATAGTATTATTTGGCTGA 696
Db      234 -----LeuProArgArgLysLeuArgSerAlaArgAlaGlyArgLeuProGly 249
QY      697 CCACAGAGAGCGCTTAATGAGAGTGAATTTTAAAGAACTTAAATACAAATCTAG 756
Db      250 -----GlyProLysTrp----- 253
QY      757 AATGATTTTCTAATATGAAAGGCGCTGCACAAAAGTAAGTATGTTCTTTATTAAGAG 816
Db      254 -----GlyAlaValAspGlyArgTyrAsnSerMetAlaLysHis 266
QY      817 GTACAAAG 825
Db      267 ValGlnGln 269

RESULT 8
US-10-369-493-19786
; Sequence 19786, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 19786
; LENGTH: 367
; TYPE: PRT
; ORGANISM: No. US20030233675A1loc punctiforme
```



```
Db 331 llelleGluSerLeuGlnProAspSerTyraenProTyrrTyserlleGlnLeuGln--- 349
Qy 61 GCAGGAGATATAGTACTGTAATACTTAAGCAAAACAGATGAGGTGATACA--GCA 117
Db 350 -----GluAenlySGlnLysAspGlnLysAsnSerGlnAla 361
Qy 118 TCATAT--GGAATTGATATGAGCTAAAGAAATATATATTAAGCTTACGCTTGCTGA 174
Db 362 ThrTyraenGlnLeuGlnIleGlnIleLysSerLeuLysSerlleLeuAsnLeuPheAsp 381
Qy 175 AGTGTATTTCAATAGTATCT-----AACTTCGCGATGACATTAATATAG 219
Db 382 PheAsnIleThrAsnLeuSerlleAspAlalleTyraenLeuAsnAspGlnLeuIleLys 401
Qy 220 AATAAGAGACTGCTGAG-----AGATTATATCAAGATAT 255
Db 402 LysPheGlnLysAlaAsnSerGlnIleAsnAspValalleLyslleLysGlnIleThr 421
Qy 256 ATGTCTAATCTTTCATCTGCACTATTAGGTGAATGTAATGTAATTTCTTAAGATATG 315
Db 422 LeuSerAsn-PheLysGlnIleTyrrLeuPheLysTyrrPheIleLysIleLysGlnIle 441
Qy 316 GCAATGTGTTTATTAAGAATGAC--TGATTTTGAAGCTCAAT----- 358
Db 441 easnIleGlnLeuAsnAsnGlnAsnAspProIleLysLysAlaAsnAspIleGlnIleThr 461
Qy 359 -----ATCTCAAAACATTT-----GGAATGTTCTCGAGCTTGAATTAACCATTG 405
Db 461 eGlnIleLysLysSerTyrrGlnLysAspLeuPheLeuAsnIleLysIleLeuLys 481
Qy 406 AGTGTCTTATTCAG-----ATGACGATTAATTTATTAAGCATATATTTTCTCTGACG 459
Db 481 sleuAlaProAsnGlnLysLeuThrValAsnAsnLeuGlnLysIleAspLeuThrTyrrAl 501
Qy 460 GAAATTCACAGTGAAGAAATCAACAATCAAAATGCCGAGATTTTAAATTAATATGAT 519
Db 501 aserAspSerLeuGlnLys-----ProGlnGln-----LeuIle 512
Qy 520 TTCTTATTTAATCTTATCTGCTGTAACTTCACTGCGAAGAGATTTTCAAAAACTTT 579
Db 512 eAlaIleTyrrLeuGln-----LeuAsnAsnAsnGlnGln--ArgGlnPheSer----- 526
Qy 580 TACAATGATTAAGAGCTAAATCATTAAGAAATTAATTAAGAGA----- 624
Db 527 -----LysGlnLeuGlnAsnLeuSerValArgValPheThrIlelle 540
Qy 625 ---AAAAAATTTCTTAACCTTTCTTTCGACCAAGCGACAGATTAACGACAGATA 681
Db 541 AspLysLysLeuTyrrIlePro-----LysAlaLysGlnLeuIleAsp----- 554
Qy 682 GGTATTTTGGCTGACCAACAGAGCGCTTAATGAGAGATGATTTTAAAGACTTAATA 741
Db 555 -----GluSerLysLysIlelleGlnIleThrLysAsnGlnIleThr 567
Qy 742 AATAACAAATCTAGAAATGATTTTCTAATATGAAAGGGCTGCAAAACAAAGTAACT 801
Db 568 AsnAsnLysAlaLys-----LeuGlnArgLeuGln-----LysPheThr 580
Qy 802 TCATTTTAAAGAGGTACAAAAGGTACGCTCCACAGACAGCGCAAAAGTATGAT 861
Db 581 SerPheValGlnGln----- 585
Qy 862 ACAGCAGATGCGATTAACCTGGAATAATTCGCGAATATTAATTAAGTGAAGCTTAAC 921
Db 586 GlnIleSerAsnGlnLysValSerSerTyrrIleAsnLeuTyrrlleGlnIleArgLeuThr 605
Qy 922 CAA 924
Db 606 Lys 606
```

```
RESULT 10
US-10-437-963-200517
; Sequence 200517, Application US/10437963

; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 200517
; LENGTH: 1162
; TYPE: PRT
; ORGANISM: Oryza sativa
; NAME/KEY: unsure
; FEATURE:
; LOCATION: (1) --(1162)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_9597C.1.pep
; US-10-437-963-200517

Alignment Scores:
Pred. No.: 1.5 Length: 1162
Score: 98.50 Matches: 63
Percent Similarity: 38.11% Conservative: 38
Best Local Similarity: 23.77% Mismatches: 100
Query Match: 5.68% Indels: 64
Db: 16 Gaps: 10

US-10-617-962-2 (1-1008) x US-10-437-963-200517 (1-1162)
Qy 172 GTAAGTGTATTCATTAATGATATCTAACTTCGATGACTATTAAGATAAGAGACT 231
Db 816 LeuSerGlnAlaIleSerLeuSer-----Asp 824
Qy 232 GCTGAGAGATTAATCAAGAAATATATGCTTAATCTTCACTGCACTTTAGCGTAAT 291
Db 825 GlyAspSerAsnTyrrGlnAspProSerSerSerValAsnSerAlaSerHisSerAspAsn 844
Qy 232 GGTATCAAAATTTCTAAGATATGCGCAATGCTTTTATTAAGATGACATG-----GAT 345
Db 845 GluLeuAsnAsnSerGlnTyrrGlySerAsnAspIleLysLeuAsnGlnAlaTyrrSerGly 864
Qy 222 GGTATCAAAATTTCTAAGATATGCGCAATGCTTTTATTAAGATGACATG-----GAT 345
Db 845 GluLeuAsnAsnSerGlnTyrrGlySerAsnAspIleLysLeuAsnGlnAlaTyrrSerGly 864
Qy 346 TTGGAAGGTCAATATCTCAAAAACATTTGGAATGTTCTGAGCTTGAATAATTAACCATTG 405
Db 865 PheAspAlaGlnAlaSerGlnIle-----GlnLysProThr 877
Qy 406 AGTGTATTCAGATGACGATTAATTAATTAAGCACTATATTTTTC---TCTGTACAGAA 462
Db 878 ArgAlaTyrrSerAspSerSerAsnMetSerSerLeuAsnIleThrPheGlnAspProAlaGly 897
Qy 443 ATTCACAGCGAGAAATCAACAAATCAAAATCCGCAATTTTAAATTAATTAATTAATTC 522
Db 898 IleProProGlnIleTyrrProThrAsnValSerValHisLeuLeuAsnLeuValAspLys 917
Qy 523 TTATTTACCTTA-----TCTGCTGAATTCACCTGGAAGAG 561
Db 918 ValPheGlnLeuLysArgArgGlyTyrrIleArgLysGlnIleValIleThrIleSerLysGln 937
Qy 562 ATTTTCAAAAACCTTTACAGATGATTAAGGCTAATATCAATTAAGAGATTAATTAATGAG 621
Db 938 IleLeuGlnLeuValMetGlnAspAlaIleAsp-----Gln-TyrrIleLeuArg 953
Qy 622 AGAAAAAATCTTTCAAAACCTTTCTTGTGACCAACGCGAGAGATTAACCGAGCGAGATA 681
Db 953 GlnIleAsnTyrrLeu-----ArgArgAspGlnValIleValGlnGln 967
```

QY 682 GGTATTGGCTGACCAACAGAGCGCTAAATGAGATGATTTTAAAGACTTAA 741
 Db 967 yllegrttrpgleinaepthleutrpProkmslly--ValphepThArgleuasp 986
 QY 742 AAT-----AACAAATCTAGAAATGATTTTCTAAT 771
 Db 967 GlyTyrgInglyAsnalglYProserGlnpheaplyshIaProserGlySerIaasn 1006
 QY 772 ATG--GAAGGCGCTGCAAAACAAAGTATGTTCTATT----- 807
 Db 1007 GlnAlaSerGlyAsnarGlyshapSerAlaSerSerPheglugInglInleuGlnAlaSer 1026
 QY 808 -----ATTAAGAGGTACAAAAGGTAAAGCTCCACAGACAGCAAA 852
 Db 1027 ArganaAlaSerGlnVallyshLeuLeuLeuGlyGlyThrProThrLeuValser 1046
 QY 853 AGTATTGCT 861
 Db 1047 llellegly 1049

RESULT 11
 US-10-282-122A-54429
 ; Sequence 54429, Application US/10282122A
 ; Publication No. US20040029129A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Wang, Liangu
 ; APPLICANT: Zamudio, Carlos
 ; APPLICANT: Malone, Cheryl
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlsen, Kari
 ; APPLICANT: Zykend, Judith
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John
 ; APPLICANT: Carr, Grant
 ; APPLICANT: Yamamoto, Robert
 ; APPLICANT: Foreyth, R.
 ; APPLICANT: Xu, H.
 ; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
 ; FILE REFERENCE: ELITRA.034A
 ; CURRENT APPLICATION NUMBER: US/10/282.122A
 ; PRIOR FILING DATE: 2003-02-20
 ; PRIOR APPLICATION NUMBER: 60/191.078
 ; PRIOR FILING DATE: 2000-03-21
 ; PRIOR APPLICATION NUMBER: 60/206.848
 ; PRIOR FILING DATE: 2000-05-23
 ; PRIOR APPLICATION NUMBER: 60/207.727
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: 60/230.335
 ; PRIOR FILING DATE: 2000-09-06
 ; PRIOR APPLICATION NUMBER: 60/230.347
 ; PRIOR FILING DATE: 2000-09-09
 ; PRIOR APPLICATION NUMBER: 60/242.578
 ; PRIOR FILING DATE: 2000-10-23
 ; PRIOR APPLICATION NUMBER: 60/253.625
 ; PRIOR FILING DATE: 2000-11-27
 ; PRIOR APPLICATION NUMBER: 60/257.931
 ; PRIOR FILING DATE: 2000-12-22
 ; PRIOR APPLICATION NUMBER: 60/267.636
 ; PRIOR FILING DATE: 2001-02-09
 ; PRIOR APPLICATION NUMBER: 60/269.308
 ; PRIOR FILING DATE: 2001-02-16
 ; Remaining Prior Application data removed - See file wrapper or PALM.
 ; NUMBER OF SEQ ID NOS: 78614
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 54429
 ; LENGTH: 580
 ; TYPE: PRT
 ; ORGANISM: Campylobacter jejuni
 US-10-282-122A-54429

Alignment Scores: 1.42 Length: 580

Score: 98.00 Matches: 74
 Percent Similarity: 35.99% Conservative: 48
 Best Local Similarity: 21.83% Mismatches: 135
 Query Match: 5.65% Indels: 82
 DB: 15 Gaps: 14

US-10-617-962-2 (1-1008) x US-10-282-122A-54429 (1-580)

QY 181 ATTGATATGATATGATTAACCTCCGATGACATATTAAGATTAAGAGACTGAGAGA 240
 Db 103 LeuysasnValleuArgleuaspMecaspPheylsArgTYrArgSerGlyGluLeu 122
 QY 241 ATTATCAAGATATATGATCTTAATCTTCACTGACATATTAGGTAAGATGATCA 300
 Db 123 ValserArg-----CysThrAsnAspIleasnIaleu-----Glnser 135
 QY 301 ATTTCTAAAGATATGACCAATATGTTTATTAAGATGAACATTTGAAAGTCATAT 360
 Db 136 IleValSerAsnIleIleProaspPhepheArgGlnleuThrAlaIleGlyLeuLeu 155
 QY 361 CCTGAAACATTTGGAAATGTTCTGAGCTTGA----- 393
 Db 156 AlaValValleuTyrgInserProThrleuAlaPhepheAlaValIleleuProleu 175
 QY 394 -----AATTAACCATTTGAGTCTTAT----- 414
 Db 176 AlaIlePheProleuValtrPheAlaIlelyshleuTyshTyAlaArgAsnIleGln 195
 QY 415 -----TCAGATGACATTAATTTATTAACATATTTTCTCTGACAGAAATT 465
 Db 196 GluthraAsnSerAspLeuSerTyreulGlyIuIlePheSerAsnIleGlyleuIle 215
 QY 466 CCACTGAGAAATCAACATCAATCAATGCCCAATGTTTAAATTAATGATTTCTTA 525
 Db 216 LysAlaAsnAspAsnGlyshLysgluSeraspIleAlaIleAsnAspThrleu 235
 QY 526 TTACCTTATCT-----CCTGAATCTCACTG----- 552
 Db 236 CysIshleuAsnleuLysSerAlaArgIleAspIleleuThrIleSerProleuMecaspMet 255
 QY 553 -----GGAAGGAGATTTTCAAAA 573
 Db 256 MetGlySerValGlyValAlaValIleIleValGlyGlyArgGlnValIle----- 273
 QY 574 AACTTTTACATGATTTAGAGCTTAATCACTTAAGATTTATTAAGAGAAAAAATT 633
 Db 274 -----Asngly-----SerMetSerValGlySerPheIleSerPheValSerAla 288
 QY 634 TCTAAACCTTCTTTCGACACCGACAGATTAACCTGATGACAGATAGTTATTGGCT 693
 Db 289 LeupheAlaIleTythrProleuLysArgLeu---SerSerleuTyrcIlyshleuGln 307
 QY 694 GGACCAACAGAGCG-----CTTAATGAGA 720
 Db 308 GlyAlaValAlaIleAsnSerGluArgThrPheTyreuleuAspLeuGlnProGlnIleLys 327
 QY 721 GTGACTTTAAAGAACTTAAATAAATPAACAAATCTAGATGATTTTCTTAATGAGAGG 780
 Db 328 glylySerLysGlnleuLysasnIleGlnu---LysIleSerPhegluAsnValIleuPhe 346
 QY 781 GCTGCAAAACAAAGTATGATTTTAAAGAGGTACAAAGGATAGCGTCCACAG 840
 Db 347 AlaTyrgIuAsnProthIshLysSerValleuTyshGlyValaAsnPheAspPheVallyshGly 366
 QY 841 ACAGCAGGAAAGTATGATTTGATACCGACAGTGCAGTAACTGGAATAATGCGCAATAT 900
 Db 367 GlnMecleuAlaIleuValGlyThr---SerGlyGlyGlyLysSerSerIleIleAsnleu 385
 QY 901 TTATATAGTGAAGCTTAAGCAAAAGACAGGGTAACCTTATCAACAAATGATCTGAC 960
 Db 386 LeuMecTyrrPheTyrgIlyshGlnLysGlyshLysIleleuLeuAsnGlnIuAspIle--- 404
 QY 961 AATCAATGACGGTTCATAGT-----GTTGAACCTCATTTATTAATAATATA 1005


```

FEATURE:
OTHER INFORMATION: Clone ID: MRT4577_17949C.1.pep
US-10-425-115-271828

Alignment Scores:
Pred. No.: 2.09 Length: 1770
Score: 97.50 Matches: 71
Percent Similarity: 36.88% Conservative: 47
Best Local Similarity: 22.19% Mismatches: 117
Query Match: 5.62% Indels: 85
DB: 17 Gaps: 16

US-10-617-962-2 (1-1008) x US-10-425-115-271828 (1-1770)

QY 103 GAGGCTCATACGATCATATGGAATGTAAT----- 135
DB 1208 GlnGlyTyrThrArgLleGlnGlyLeuAspPheGlyGluArgPheAlaProValAlaArg 1227
QY 136 ---CGAGCTAAGAAATATAATTAGCTTACGCTTGCTGTAAGTGTATTCAATGTA 192
DB 1228 LeuGluAlaIleArgLleLeuLeuAlaTyrAlaCysAla-----HisAsnIle 1243
QY 193 TCTAACTCTCGATGACTAT-----TATAAGAAATAAGAGACTGCTGAGAAAT 243
DB 1244 LysLeuTyrGlnMetAspValLysSerAlaPheLeuAsnGlyLysIleSerGluLeuVal 1263
QY 244 TATCAAGAA-----TATATGCTTAATCTT 267
DB 1264 TyrValGlnGlnProProAspPheGluAspProLysArgProIleHisValPheLysLeu 1283
QY 268 TCATCTGCACTATTAGTGAATAATGCTGATCAAAATTTCTTAAGATATGCAAAATGCTTT 327
DB 1284 SerLysAlaLeuTyrGly-----LeuLysGlnAlaProArgAlaTrp 1297
QY 328 TATAGAAATGAATGATTTTGAAGTCAATATCTCAAAACATTGGAAATGTT----- 381
DB 1298 TyrGluArgLeuArgAspPheLeuLeuSerLysAspPheArgLleGlyLysValAspThr 1317
QY 382 -----CCTGAGCTTGAATAATTAACCATTTG-----AGTCTTATTCAGATGAC 423
DB 1318 ThrLeuPheThrLysIleIleGlyLysAspLeuPheValCysGlnIleTyrValAspAsp 1337
QY 424 GATAAATATTAGACTATATTTTCTCTGACAGAAATCCATGAGAGAAATGCA 483
DB 1338 -----IleIlePheGlySerThrAsnGlnLeuPheCysGlnIle----- 1350
QY 484 CAATCAAAATGCCGAAGAATTTTAAATTAATTAATGATTTCTTAATCTTATCTGCTGA 543
DB 1351 -----PheGlyLysMetMetSerLysGluPheGluMetSerMetIle 1364
QY 544 ACTTACTGGGAGAGAGATTTTTCAAAAAATTTCATCATGATGATTAGAGCTTAATCA 603
DB 1365 GlyIleLeu-----SerPhePheLeuGlnIleLeuGlnIleLysGln 1377
QY 604 TTAGAGAAAT-----TATATTGAGAGAAAAAACTTTCAACCTTTCTTTCGA----- 651
DB 1378 LeuLysAspGlyLlePheIleSerGlnSerLysTyrLeuLysAspMetLeuLysPhe 1397
QY 652 -----CCACCGCAGAGATTACTGATGCGAGAAATGAGTTATTTG----- 690
DB 1398 GlyLeuGluAsnAlaLysProLleLysThrProMetAlaThrAsnGlnHisLeuAspLeu 1417
QY 691 -----GCTGACCAACAGAA-----GGCCTTAATGAGAGCTGAGTTT 729
DB 1418 AspGlnGlyAlaSerProArgGlnIleHisLeuLysAlaAlaLysArgLleLeuArgTyr 1437
QY 720 AAAAGAACTTAATAATACAAATCTAGAAATGATTTTCTAATATGAGAGGCTGCAAAA 789
DB 1438 LeuLysTyrThrProAsnIleGlyLeuLysGlyLeuGlyLysLeuArgLysLeuAla 1457
QY 790 CAAAAGTATAGTTCAATTATTAAGAAGGTACAAAAGGTAAC--GCTCCACAGACAGCA 846
DB 1458 GlyArgPheLysSerLysArgProArgValAspAspAlaAspTyrAsnProThrThrAsp 1477

```

```

QY 847 GCGAAAGTATTTGTTAGACCCAGTGGCAGT-----AACCTGAAAAAATGCCGAAT 897
DB 1478 SerGluAlaGlnSerLeuAlaGlyGlySerValSerMetAspThrGlnAspValProHis 1497

RESULT 14
US-10-425-115-249308
; Sequence 249308, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCES: 38-21(53222)B
; CURRENT FILING DATE: US/10/425, 115
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 249308
; LENGTH: 878
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_158956C.1.pep
US-10-425-115-249308

Alignment Scores:
Pred. No.: 1.97 Length: 878
Score: 97.00 Matches: 74
Percent Similarity: 34.75% Conservative: 49
Best Local Similarity: 20.90% Mismatches: 137
Query Match: 5.59% Indels: 94
DB: 17 Gaps: 14

US-10-617-962-2 (1-1008) x US-10-425-115-249308 (1-878)

QY 19 CCTGATGATAGAGTATATCCACCGCTTGAAGAAATAGCAGAGATATAGTACT 78
DB 523 ProAspIleLysThrSer--SerProIleGlnLysGlnGlyAlaGlnLeuTyrThrArg 541
QY 79 -----ATATCAAACTTTAAACAA----- 96
DB 542 SerMetPheLeuLysPheGlnGlnLeuIleCysAlaThrAlaTyrThrAlaGluMet 561
QY 97 ---ACAGATGAGGTCATACGATCATATGGAAT----- 129
DB 562 ValLysGlnGlnGlyAsnAlaSerIleTyrThrValAlaIleSerGlnGlySerGluLys 581
QY 129 ----- 129
DB 582 SerValThrValGlnPheCysSerSerGlySerSerAlaThrCysSerCysArgMetPhe 601
QY 130 GAATATGAGGTAAGAAATATATATAGCTTACGCTTGCTGTAAGTGTATTGATTAAT 189
DB 602 GlnTyrPheGly-----IleValCysArgHisIleLeuThrValPheGlyValArgGly 619
QY 190 GTATCTAAACTCTCGATGACTATATTAAGATTAAGAAAGAACTGCTGAGAAATTTATCA 249
DB 620 ValSerAlaLeuProSerHisIleTyrPheValLysArgTyrThr----- 633
QY 250 GAATATATGTTAATCTTTCAATCGCAATATTTAGGTGAATAATGGTATCAATTTCTAAA 309
DB 634 -----LysAsnAlaLeuAspArgSerLeuGlnLysAsnValAspGlnValSerArg 650
QY 310 GATATGCAAAATGCTTTTATTAAGATAGAACTGATTTTGAAGGTCATATCTCAAAAC 369
DB 651 ValGlnProLysGlnGlnGlnArgSerSerAlaGluAspGlyGlnGlnSerPro----- 668
QY 370 ATTGGAATGTTCTGACCTTGAATAATCAACATTGAGCTTATTCAGATGACGATAAA 429
DB 669 ThrTrpArgTyrAsnSerLeuCysHisGlnAlaLeuArgTyrAlaGlnGlnGlyAlaSer 688

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_nzp model

Run on: November 21, 2004, 07:40:10 ; Search time 92 Seconds

(without alignments)
7860.851 Million cell updates/sec

Title: US-10-617-962-2
Perfect score: 1735
Sequence: 1 atgcttatacattacacc.....ctcattataaataatgta 1008

Scoring table:
BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 2002273 seqs, 35872929 residues

Total number of hits satisfying chosen parameters: 4004546

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+nzp.model -DEV=xlp
-O=/cgn2.1/USPTO_epool_p/US10617962/runat.17112004.163921.9743/app_query.fasta_1.1159
-DB=A Geneseq 23Sep04 -QFWT=fastan -SUFFIX=rag -MINMATCH=0.1 -LOOPCL=0
-LOOEXT=0 -UNITS=bites -START=1 -END=1 -MATRIX=blowum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pro -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10617962 @CGN 1.1.154 @runat.17112004.163921.9743 -NCRP=6 -ICPU=3
-NO MMAP -LARGESQUERY -NEG SCORES=0 -WAIT -DSPELOCK=100 -LONGLOC
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
1: A_Geneseq_23Sep04:*
2: geneseqp19808:*
3: geneseqp19908:*
4: geneseqp20018:*
5: geneseqp20028:*
6: geneseqp20038:*
7: geneseqp20038:*
8: geneseqp20048:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1734	99.9	335	2	AAW97811 Photorhab
2	1716	98.9	340	6	ABM67356 Photorhab
3	1276	73.5	368	2	AAW97810 Xenorhabd
4	940.5	54.2	278	2	AAW68961 Xenorhabd
5	108.5	6.3	1654	1	AAW50777 Sequence
6	108	6.2	940	5	ABW47334 Listeria
7	105.5	6.1	816	6	ADA35390 Acinetoba
8	105.5	6.1	1270	7	ADW83729 Bacillus
9	104.5	6.0	2197	8	ADK16573 Nanorarcha
10	104	6.0	583	6	ABW63439 S suis ar

11	101.5	5.9	1639	2	AAW54145
12	101.5	5.9	1639	5	AAE29345
13	101	5.8	566	6	ABR53129
14	101	5.8	566	7	ADK62560
15	100	5.8	2404	8	ADP25426
16	98.5	5.7	616	6	ABU48940
17	98	5.6	580	6	ABU26505
18	98	5.6	954	6	ABU35684
19	98	5.6	1005	5	ABU35684
20	98	5.6	1413	4	ABW62045
21	97.5	5.6	511	2	AAW37593
22	97	5.6	1714	3	AAW18275
23	96.5	5.6	685	4	AAU37884
24	96.5	5.6	685	6	ABU02144
25	96.5	5.6	685	6	ABP81632
26	96.5	5.6	685	6	ABU46194
27	96	5.5	384	7	ADCI9754
28	96	5.5	512	4	ABW71449
29	96	5.5	623	6	ABR53842
30	96	5.5	623	7	ADK64956
31	96	5.5	645	2	AAW86966
32	96	5.5	879	6	ABU26364
33	96	5.5	912	2	AAW94029
34	96	5.5	3168	7	ABO67550
35	95.5	5.5	303	6	ABU23912
36	95.5	5.5	360	6	ABU26457
37	95.5	5.5	680	8	ADK46676
38	95.5	5.5	915	2	AAW94700
39	95.5	5.5	1791	8	ADP25441
40	95	5.5	500	5	ABP29236
41	95	5.5	577	4	ABG05216
42	95	5.5	577	4	ABG15756
43	95	5.5	577	4	ABG29786
44	95	5.5	732	4	ABG13490
45	95	5.5	1048	4	ABG21867

ALIGNMENTS

RESULT 1	AAW97811	standard; protein; 335 AA.
AC	AAW97811;	
DT	21-MAY-1999	(first entry)
DE	Photorhabdus luminescens insecticidal toxin protein PIV6tox1.	
KW	Toxin; PIV6tox1; biological control; insecticide.	
OS	Photorhabdus luminescens.	
PN	W09903328-A1.	
PD	28-JAN-1999.	
PF	17-JUL-1998;	98WO-AU000562.
PR	17-JUL-1997;	97AU-00008088.
PA	(CSIR) COMMONWEALTH SCI & IND RES ORG.	
FI	Eaet PD;	
DR	WPI; 1999-131733/11.	
DR	N-PsDB; AAX07350; AAX07352.	
PT	New insecticidal toxin genes - extracted from Xenorhabdus nematophilus	
PT	A24 and Photorhabdus luminescens.	
XX	Claim 6; Page 29-30; 48pp; English.	

CC This is the amino acid sequence of the insecticidally-active toxin
 CC protein, termed P1v6toxi, of Photobhabdus luminescens strain V16/1. It
 CC was deduced from the cDNA region (see AAX07350) of the P1v6toxi gene,
 CC which was isolated from a DNA library using the Xenorhabdus nematophilus
 CC toxB4 gene (see AAX07349) as probe. The invention relates to the
 CC identification and isolation of polynucleotide molecules encoding a new
 CC class of protein insecticidal toxins which are produced by bacteria of
 CC the genera Xenorhabdus and Photobhabdus. The polynucleotide molecules may
 CC be incorporated e.g. into insect-specific viruses (e.g. entomopox and
 CC nuclear polyhedrosis viruses), bacteria (e.g. Gracilicutes, Firmicutes,
 CC Tenericutes and Mendosicutes), protozoa, yeast and plants (particularly
 CC cereals such as wheat and barley, vegetables such as tomato and potato,
 CC fruit trees such as citrus and apple, tobacco and cotton) for control of
 CC pest insects

CC Sequence 335 AA;

CC Alignment Scores:

Pred. No.:	1,226-168	Length:	335
Score:	1734.00	Matches:	335
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	99.94%	Indels:	0
DB:	2	Gaps:	0

US-10-617-962-2 (1-1008) x AAW97811 (1-335)

QY 1 ATGGTTATACATTAATTAACCTGATGATAGAGTGTATCCACCCGTTGAAAAAGCAATA 60
 DB 1 MetValIleGlnLeuThrProAspAspArgSerGlyTyrProProValGluLysGlnIle 20
 QY 61 GCAGAGATATATGTAACGTAATTAACCTTAAGCAACAGATGAGGCTCATACAGATCA 120
 DB 21 AlaGlyAspIleValAlaArgIleLeuAsnPheLysGlnThrAspGlnGlyHisThrAlaSer 40
 QY 121 TATGGAATTAATGATCGAGCTAAGAAATAATATTAGCTTATGGCTGTAAAGTGT 180
 DB 41 TyrGlyIleGlnTyrArgAlaLysLysIleIleLeuAlaTyrAlaLeuAlaValSerGly 60
 QY 181 ATTCATAATGATTAACCTTCCTGATGACTATTATAGAAATAAGAGACCTGAGAGA 240
 DB 61 IleHisAsnValSerIleLeuProAspAspArgTyrTyrLysAsnLysGlnThrAlaGlnArg 80
 QY 241 ATTATTAAGAAATATATGCTTAATCTTTCATCTGCATATTAGTGAAATAAGTGATCAA 300
 DB 81 IleTyrGlnGlnLysTyrMetSerAsnLeuSerSerAlaLeuLeuGlyGlnLysGlnArg 100
 QY 301 ATTTCATAAGATATGCGAAATGCTTTTATAGAAATGAGATTTTGAAGTCAATAT 360
 DB 101 IleSerLysAspMetCalaAsnGlyPheTyrLysAsnGlnLeuAspPheGlnGlyGlnTyr 120
 QY 361 CCTCAAAACATTGGATGATGCTTCCTGAGCTTGAATAAATCAATTGAGTCTTATTCAGAT 420
 DB 121 ProGlnAsnIleIleTyrAsnValProGlnLeuGlnLysLeuSerIleTyrSerAsp 140
 QY 421 GAGCAATAATTAATTAAGCACTATATTTTCTCTGACAGCAAAATTCACATGAGGAAAT 480
 DB 141 AspAspLysLysLeuLeuAlaLeuTyrPhePheSerValGlnGlnIleProLeuGlnGln 160
 QY 481 CAACAAATCAATGCGCGAGATTTTATTAATTAATGATTTCTTATTAATCTTATTCGT 540
 DB 161 GlnGlnSerAsnAlaIleArgPhePheLysLeuIleAspPheLeuPheThrLeuSerAla 180
 QY 541 GTAATCTCACTGGAAGAGAGATTTTTCATTAATAAATTTTACATGATGATGAGCTTAA 600
 DB 181 ValThrSerLeuGlyArgArgIlePheSerLysAsnPheTyrAsnGlyLeuGlnAlaLys 200
 QY 601 TCATTAGAGATTAATTAATGAGAAAAAATCTTCTTAACTTTCTTTCGACACCGCAG 660
 DB 201 SerLeuGlnLysAsnTyrIleGlnArgLysLysLeuSerLysProPhePheArgProProGln 220
 QY 661 AGATTACCTGATGCGAGATAGTTATTTGGCTGACCAACAGAACCGCTTAATGAGAA 720

DB 221 ArgLeuProAspArgLysArgIleGlyTyrIleuAlaGlyProThrGlnAlaProLysTyrArg 240
 QY 721 GTGACTTTTAAAGACTTAAATAATCAATCTAGAAATGATTTTCTTAATTAATGAGACG 780
 DB 241 ValSerPheLysGlnLeuLysAsnLysSerIleArgAsnGlyPheSerAsnMetGlnGly 260
 QY 781 GGTGCAAAACAAAGATATGATTTTATTTTAAAGGATACAAAGGATTAACCTTCACAG 840
 DB 261 AlaAlaLysGlnLysTyrSerSerPheIleLysGlnValGlnLysGlyAsnAlaProGln 280
 QY 841 ACAGCAGCAAAAAGATTAATGATGATACAGCAGTGACAGTAACTGCAAAAATTCGCCAATAT 900
 DB 281 ThrAlaAlaLysSerIleGlyThrAlaSerGlySerAsnLeuGlnLysLeuProAsnAsn 300
 QY 901 TTATATAGTGTAGGCTTAAGCCAAAAGACAGGATTAACCTTACTCAAAATGATATGAC 960
 DB 301 LeuTyrSerValArgLeuSerSerLysAspArgValThrPheThrGlnAsnAspThrAsp 320
 QY 961 AATACAAATGACGCTTCAATGATGTTGGAATCTCATTTTAAATAATTA 1005
 DB 321 AsnThrMetThrValHisSerValGlyThrHisTyrLysAsnIle 335
 RESULT 2
 ABM67356
 ID ABM67356 standard; protein, 340 AA.
 AC ABM67356;
 DT 20-NOV-2003 (first entry)
 DE Photobhabdus luminescens protein sequence #453.
 KW Antibacterial, fungicide, insecticide, polymorphism; genetic analysis;
 KW detection; food; gene expression; plant; animal; microorganism; toxin;
 KW antibiotic; bioplastic; virulence factor; disease model; plague;
 KW whooping cough.
 OS Photobhabdus luminescens.
 PN WO200294867-A2.
 PD 28-NOV-2002.
 PF 07-FEB-2002; 2002WO-1B003040.
 PR 07-FEB-2001; 2001FR-00001659.
 PA (INSP) INST PASTEUR.
 PA (CNRS) CNRS CENT NAT RECH SCI.
 PI Duchaud E, Taourit S, Glaser P, Frangeul L, Kunst F, Danchin A;
 PI Buchriesser C;
 DR MPI; 2003-148459/14.
 PS Claim 2; SEQ ID NO 453; 1205bp; French.
 PT Genomic sequence of Photobhabdus luminescens and encoded polypeptides,
 PT useful e.g. as therapeutic antimicrobials and agricultural pesticides.
 CC The invention relates to the isolation of genes and their encoded
 CC proteins from Photobhabdus luminescens. The isolated sequences are
 CC sources of probes and primers for detecting the genome of P. luminescens
 CC and related species; to study polymorphisms; for gene analysis and for
 CC detection/amplification of the genes. Antibodies (Ab) raised against the
 CC polypeptides encoded by the genes are used for detection/identification
 CC of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that
 CC carry a gene-containing vector are used to select compounds that
 CC modulate, regulate, induce or inhibit expression of the genes in plants,
 CC animals or microorganisms other than P. luminescens and are able to alter
 CC response or sensitivity to toxins and antibiotics produced by P.
 CC luminescens. Cells transformed to express the genes are useful for
 CC recombinant production of the proteins, particularly toxins and

CC antibacterials useful as insecticides, bactericides and fungicides. The
 CC genes, proteins, vectors containing the genes and Ab are also useful
 CC therapeutically (to treat microbial infection by bacteria or fungi that
 CC are sensitive to *P. luminescens*-encoded toxins or antibiotics) and as
 CC biopesticides. Other uses of the genes and the proteins are as virulence
 CC factors and for identifying targets of human diseases for which *P.*
 CC *luminescens* is a model (particularly plague and whooping cough). This
 CC sequence represents one of the isolated *P. luminescens* proteins
 XX
 SO Sequence 340 AA;

Alignment Scores:

Pred. No.: 8,59e-167 Length: 340
 Score: 1716.00 Matches: 331
 Percent Similarity: 99.10% Conservative: 1
 Best Local Similarity: 98.81% Mismatches: 3
 Query Match: 98.90% Indels: 0
 DB: Gaps: 0

US-10-617-962-2 (1-1008) x ABM67356 (1-340)

QY 1 ATGGTTATCAATTAACCTGATGATGAAAGTGATATCCACCCGTTGAAAGCAATA 60
 DB 5 MetValIleGlnLeuThrProAspAspArgSerGlyTyrProProValGlnLeuGlnIle 24
 QY 61 GCAGGAGATATAGTACGTATCTAACTTAAAGCAAGATGAGGTCATACAGCATCA 120
 DB 25 AlaGlyAspIleValAlaArgIleLeuAsnPhelyGlnThrAspGlyHisThrAlaSer 44
 QY 121 TATGGAATGATATCGAGCTAAGAAATAATATTAGCTTACGCTTGGCTGAAGTGT 180
 DB 45 TyrGlyIleGlnTyrArgAlaLysIleIleLeuAlaTyrAlaLeuAlaValSerGly 64
 QY 181 ATTCATATGATCTAACTTCTGATGATCTTTTAAAGATAAAGAGACTGCTGAGCA 240
 DB 65 IleHisAsnValSerTyrLeuProAspAspTyrTyrAsnLysGlnThrAlaGlnArg 84
 QY 241 ATTTATCAAGATATATGATCTTATCTTATCATCTGACATTTAGGGAATGGTATCAA 300
 DB 85 IleTyrGlnGlnTyrMetSerAsnLeuSerAlaLeuLeuGlnGlnAsnLysAspGln 104
 QY 301 ATTTCTAAAGATATGCAAAATGTTTATTAAGATGAATCTGATTTTGAAGTCAATAT 360
 DB 105 IleSerTyrAspMetAlaAsnGlyPheTyrLysAsnGlnLysLeuAspPheGlnGlyLysTyr 124
 QY 361 CCTCAAAACATTTGGAATGTTCTTGAAGCTTGAATAAACCATTGAGTCTTATTCAGAT 420
 DB 125 ProGlnAsnIleTyrAsnValProGlnLeuGlnAsnLysProLeuSerAlaTyrSerAsp 144
 QY 421 GACGATTAATTAATTAATGCACTAATTTTCTCTGACGGAATCCCTGAGGAAT 480
 DB 145 AspAspLysLeuLeuAlaLeuTyrPhePheSerValGlnGlnIleProLeuGlnLysAsn 164
 QY 481 CAACATCAATATGCGCAAGATTTTAAATTAATTAATTTCTTATTAACCTTATCTGCT 540
 DB 165 GlnGlnSerAsnAlaAlaArgPhePheLysLeuIleAspPheLeuLeuThrLeuSerAla 184
 QY 541 GTAATTCACTGCGGAGAGAGATTTTTCAAAAAATTTTACATGATGATGAGGCTTAA 600
 DB 185 ValThrSerLeuGlyArgArgIlePheSerLysAsnPhelyAsnGlyLeuGlnLysSer 204
 QY 601 TCATTAGGAATTAATTAATGAGAGAAAAAATCTTAAACCTTCTTTCGACACCGGAG 660
 DB 205 SerLeuGlnLysTyrIleGlnTyrGlyLysPheProLysPhePhePheArgProGln 224
 QY 661 AGATTACCTGATGAGAGATAGATTATTTGCTGAGCAACAGAGCGCTTAATGAGAGA 720
 DB 225 ArgLeuProAspGlyArgIleGlyTyrLeuAlaGlyProThrGlnAlaProLysTyrArg 244
 QY 721 GTGAGTTTAAAGACTTAAATTAACCAATCTAGAGATGATTTTCTTAATTTGAGAGG 780
 DB 245 ValSerPheLysGlnLeuLysAsnLysSerTyrArgAsnGlyPheSerAsnMetGlnGly 264

QY 781 GCTGCAGAAACAAAGATATAGTTCTTATTAAGAGGTACAAAAGGTACGCTCCACAG 840
 DB 265 AlaAlaLysGlnLysTyrSerSerPheIleGlyValGlnLysGlyAsnAlaProGln 284
 QY 841 ACAGCAGGAGAAAGATTTGTTGATCAGCCAGTGGAGTAACTCGGAAAAATTCGCAATAT 900
 DB 285 ThrAlaAlaLysSerIleGlyThrAlaSerIleSerAsnLeuGlnLysLeuProAsnAsn 304
 QY 901 TTATATGTGTGAGGCTTAAGCAAAAAGACAGGTTAACCCTTACTGAAATGATPACTGAC 960
 DB 305 LeuTyrSerValArgLeuSerGlnLysAspArgValThrPheThrGlnAsnAspThrAsp 324
 QY 961 AATACAAATGACGTTTCATAGTTTGAACCTCATTAATAAATAATA 1005
 DB 325 AsnThrMetThrValHisSerValGlyThrHisTyrLysAsnIle 339
 RESULT 3
 AAW97810
 ID AAW97810 standard; protein; 368 AA.
 AC AAW97810;
 XX
 AC AAW97810;
 XX
 DT 17-OCT-2003 (revised)
 DT 21-MAY-1999 (first entry)
 XX
 DE Xenorhabdus nematophilus insecticidal toxin protein toxh4.
 XX
 KM Toxin; toxh4; biological control; insecticide.
 OS Xenorhabdus nematophila.
 PN W09903328-A1.
 PD 28-JAN-1999.
 XX
 PD 28-JAN-1999.
 XX
 PF 17-JUL-1998; 98WO-AU000562.
 XX
 PR 17-JUL-1997; 97AU-00008088.
 XX
 PA (CSTR) COMMONWEALTH SCI & IND RES ORG.
 XX
 PI East PD;
 PI MPI; 1999-131733/11.
 DR N-PSDB; AAX07349; AAX07351.
 XX
 PT New insecticidal toxin genes - extracted from *Xenorhabdus nematophilus*
 PT A24 and *Photorhabdus luminescens*.
 XX
 PS Claim 6; Page 26; 48pp; English.
 XX
 CC This is the amino acid sequence of the insecticidally-active toxin
 CC protein, termed toxh4, of *Xenorhabdus nematophilus* strain A24. It was
 CC deduced from the coding region (see AAX07349) of the toxh4 gene, which
 CC was isolated from a DNA library by functional screening using *Galleria*
 CC mellonella fourth instar larvae. The invention relates to the
 CC identification and isolation of polynucleotide molecules encoding a new
 CC class of protein insecticidal toxins which are produced by bacteria of
 CC the genera *Xenorhabdus* and *Photorhabdus* (see also AAX07350). The
 CC polynucleotide molecules may be incorporated e.g. into insect-specific
 CC viruses (e.g. entomopox and nuclear polyhedrosis viruses), bacteria (e.g.
 CC *Grieciliacutes*, Firmicutes, Tenacitutes and Mendosicutes), protozoa, yeast
 CC and plants (particularly cereals such as wheat and barley, vegetables
 CC such as tomato and potato, fruit trees such as citrus and apple, tobacco
 CC and cotton) for control of pest insects. (Updated on 17-OCT-2003 to
 CC standardise OS field)
 XX
 SO Sequence 368 AA;

Alignment Scores:
 Pred. No.: 1.15e-121 Length: 368
 Score: 1276.00 Matches: 251
 Percent Similarity: 80.17% Conservative: 36

Best Local Similarity: 70.11% Mismatches: 47
 Query Match: 73.54% Indels: 24
 DB: 2 Gaps: 3
 US-10-617-962-2 (1-1008) x AAR697810 (1-368)

```

QY 4 GTTATACAAATTAAACCTGATGATAGA-----AGTGGATATTCACCC 45
  |||||
DB 11 ValIleGlnLeuThrProAspAspArgValThrProAspAspLysGlyGluTyrGlnPro 30
  |||||
QY 46 GTTGAAGAACAATAGCAGAGATATAGTACGTATTAACCTTAAGCAAAACAGATGAG 105
  |||||
DB 31 ValGluLysGlnIleAlaGlyAspIleIleArgValLeuGlnPheLysGlnThrAsnGln 50
  |||||
QY 106 GGTCTATACGATCATATGAAATTGAAATTCGAGCTAAGAAAATATATTACCTTAACGCT 165
  |||||
DB 51 SerHisThrGlyLeuTyrGlyIleAlaTyrArgAlaLysLysValIleIleAlaTyrAla 70
  |||||
QY 166 TTGGCTGTTAGTGTATTTCTATATGTTAACTTCTGATGACTTATTTAAGAAATAA 225
  |||||
DB 71 LeuAlaValSerGlyIleHisAsnValSerGlnLeuProGluAspTyrTyrLysAsnLys 90
  |||||
QY 226 GAGACTGCTGAGAAATTTATCAAGAATATATGTTCAATCTTCAATCTGACATTTAGT 285
  |||||
DB 91 AspaenthrglyArgGlyIleTyrGlnGluTyrMetSerAsnLeuLeuSerAlaLeuLeuGly 110
  |||||
QY 286 GAAAATGCGTATCAAAATTTCTAAAGATATGCAAAATGTTTATTAAGAATGAATGAT 345
  |||||
DB 111 GluAsnGlyLysAspGlnIleSerLysAspMetAlaAsnAspPheThrGlnAsnGlnLeuGln 130
  |||||
QY 346 TTTGAAGGTCATATCCCTCAAAAACATTTGGAAATGTTCCCTGAGCTTGAATAAATCAATG 405
  |||||
DB 131 PheGlyGlyGlnArgLeuLysAsnThrTyrAspIleProAspLeuGlnLysLeuLeu 150
  |||||
QY 406 AGTGCCTTATGATGACGATTAATATATATAGCACTATATTTTCTGTCACAGAAAT 465
  |||||
DB 151 GluAspTyrSerAspGlnLysLeuLeuAlaLeuTyrPhePheAlaSerGlnGlnLeu 170
  |||||
QY 466 CCACTGAGAGAAATCAACAATCAAAATGCGCAAGATTTTAAATTAATTAATGATTTCTTA 525
  |||||
DB 171 PrometGluAlaAsnGlnGlnSerAsnAlaAlaAsnPheLysValIleAspPheLeu 190
  |||||
QY 526 TTTACTTATCTGCTGATCTTCACTGCGGAGAGAGATTTTTCAAAAACCTTTTACAT 585
  |||||
DB 191 LeuIleLeuSerAlaValThrSerLeuGlyLysArgIlePheSerLysAsnPheTyrAsn 210
  |||||
QY 586 GGATTAAGAGGCTAAATCATTTAGGAATTTATTTAGAGAAAAAACTTCTTAACCTTTC 645
  |||||
DB 211 GlyLeuGlnThrLysSerLeuGlnLeuMetTyrIleGluArgLysLysLeuSerLysProPhe 230
  |||||
QY 646 TTTGACCAACCGCAGAGATTACCTGATGCGAGAAATGATTATTGCTGACCAACAGAA 705
  |||||
DB 231 PheArgProProGlnLysLeuProAspGlyArgThrGlyTyrLeuAlaGlyProThrLys 250
  |||||
QY 706 GGGCCTAAA----- 714
  |||||
DB 251 AlaProLysLeuProThrThrSerSerThrAlaThrThrSerThrAlaIleSerSerAsn 270
  |||||
QY 715 TGGAGAGTGAATTTAAAGAACTTAAAAAATAAACAATCTTAGAGATGATTTTCTAATTAAG 774
  |||||
DB 271 ThrPheGlnValSerLeuGlnLysLeuAsnArgAspAsnProSerArgAsnThrPheMetLysMet 290
  |||||
QY 775 GAAGGGGCTGCAAAAACAAAAGTATAGTTCATTTATTAAGAGATACAAAAGGTTACGCT 834
  |||||
DB 291 AspaAspAlaIleAlaLysArgLysTyrSerSerPheIleIleGlyValGlnLysGlyAsnAsp 310
  |||||
QY 835 CCACAGACAGCGCAAAAGTATGTTGTAACGACAGTGGCAATGTAACCTGCAAAAATTTGCCG 894
  |||||
DB 311 ProAlaGlnAlaIleAlaIleSerIleGlyThrLysSerGlySerAsnAsnHeGlnLysLeuGln 330
  |||||
QY 895 ---AATATTTTATATGTTGAGGCTAAGCAAAAACAGACAGTAACTTTACCAAAAT 951
  |||||
DB 331 GlyArgAspLeuTyrSerIleArgLeuSerGlnGlnHisArgValThrPheSerIleAsn 350
  |||||

```

QY 952 GATACCTGACAAATACATGACGTTCAATGCTTGGAACTCATTTATTAATAATA 1005
 ::|||
 DB 351 AsnThrAspGlnIleMetGluIleGlnSerValGlyThrHisTyrGlnAsnIle 368

RESULT 4

ID AAR68961 standard; protein; 278 AA.

XX AAR68961;

DT 12-FEB-2004 (revised)

DT 27-AUG-2003 (revised)

DT 09-AUG-1995 (first entry)

DE Xenorhabdus nematophilus strain A24 insecticidal toxin.

KM Bacterium; insecticide; pesticide; toxin.

XX Xenorhabdus nematophila.

PN WO9500647-A1.

XX 05-JAN-1995.

PF 24-JUN-1994; 94WO-AU000348.

XX 25-JUN-1993; 93AU-00009638.

PA (CSIR) COMMONWEALTH SCI & IND RES ORG.

PI Smigielski AJ, Akhurst RJ;

DR WPI; 1995-052084/07.

XX N-PSDB; AAG80839.

PT Polynucleotide(s) that encode insecticidal toxins from *Xenorhabdus* -

PT useful for the control of insect pests in the agricultural, aquatic and

PS forest industries.

XX Claim 8; Page 9; 15pp; English.

CC This insecticidal toxin which is useful for genetically engineering a

CC wide range of biological systems which will become more useful for the

CC control of insect pests detrimental to agricultural, aquatic and forest

CC industries. The protein may be expressed in recombinant organisms.

CC (updated on 25-MAR-2003 to correct PN field.) (updated on 27-AUG-2003 to

XX correct OS field.)

SQ Sequence 278 AA;

Alignment Scores:

Pred. No.: 2,666-87 Length: 278

Score: 940.50 Matches: 191

Percent Similarity: 80.22% Conservative: 24

Best Local Similarity: 71.27% Mismatches: 29

Query Match: 54.21% Indels: 25

DB: 2 Gaps: 2

```

QY 4 GTTATACAAATTAAACCTGATGATAGA-----AGTGGATATTCACCC 45
  |||||
DB 11 ValIleGlnLeuThrProAspAspArgValThrProAspAspLysGlyGluTyrGlnPro 30
  |||||
QY 46 GTTGAAGAACAATAGCAGAGATATAGTACGTATTAACCTTAAGCAAAACAGATGAG 105
  |||||
DB 31 ValGluLysGlnIleAlaGlyAspIleIleArgValLeuGlnPheLysGlnThrAsnGln 50
  |||||
QY 106 GGTCTATACGATCATATGAAATTGAAATTCGAGCTAAGAAAATATATTACCTTAACGCT 165
  |||||
DB 51 SerHisThrGlyLeuTyrGlyIleAlaTyrArgAlaLysLysValIleIleAlaTyrAla 70
  |||||

```



```
Db 425 SerpHeGlyAspLeuIleAsnProHeAspTyrThrLysGluProSerLysAsnIleTyr 444
Qy 583 AATGATTAGAGCTTAATCATTAAGAAATTATATGAGAGAAA--AACTTTCTAAA 639
Db 445 ThrAspAngluAryLysLysHeIleAsnGluIleLysGluLysIleLysIleGluLys 464
Qy 640 CCTTCTTCGACCAACCGAGATTAAGCTGATGCGAATAGTTATTTGGCTGGACCA 699
Db 465 LysLysIleGluSerAspLysSerTyrGluAspAspSerLysSerLeuAsnAspIle 484
Qy 700 ACAGAAAGGCGCTTAATGAGAGTGAAGTTTAAAGAACTTAAATGAACAACTAGCAT 759
Db 485 ThrLys-----TTTCTAATGAGAGGCTGCAAAACAAAGTATGTTTATTATTA 813
Qy 760 GGA-----TTTCTAATGAGAGGCTGCAAAACAAAGTATGTTTATTATTA 813
Db 503 AsnIleAspLeuThrAsnHeGluLysMetMetGlyLysArgTyrSer-----Tyr 519
Qy 814 GAGGTACAAAGGTTAAGCGCTCCACAGACAGACGCAAAAGTATGTTGACAGCCAGTGC 873
Db 520 LysValGluLysLeuThrHisProAsnThrPheAlaSerTyr-----GluAsnSerLys 537
Qy 874 AGTAACTGGAAGAAATTTGCGGAATATTA-----TATAGTGGAGGCTTA 918
Db 538 HisAsnLeuGluLysLeuThrLysAlaLeuLysTyrMetGluAspTyrSerLeuArg--- 556
Qy 919 AGCCAAAAAGACAGGTTAAGCTTTACTCAAAATGATACTGACATATGACGTTTCAT 978
Db 557 -----AsnIleValValGlu 561
Qy 979 AGTGTGAACTCATTTATTAATATATA 1005
Db 562 LysGluLeuLysTyrTyrLysAsnLeu 570

RESULT 6
ABB47334
ID ABB47334 standard, protein, 940 AA.
AC ABB47334;
DT 05-FEB-2002 (first entry)
DE Listeria monocytogenes protein #38.
KW Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;
  vitamin B12; bacterial infection; disease.
OS Listeria monocytogenes.
PN WO200177335-A2.
PD 18-OCT-2001.
PF 11-APR-2001; 2001WO-FR001118.
PR 11-APR-2000; 2000FR-00004629.
XX (INSP ) INST PASTEUR.
PA
XX Buchriesser C, Frangoul L, Couve E, Ruenick C, Feini H, Dehoux P,
  Duesuget O, Chetouani F, Nedjari H, Glaeser P, Kunst F, Cossart P,
  Daniels J, Goebel W, Kreft J, Kuhn M, Ng B, Vazquez-Boland JA,
  Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A,
  Chakraborty T, Dommann E, Hain T, Berche P, Charbit A, Durant L,
  Perez-Diaz J, Baguero F, Garcia Del Portillo F, Gomez-Lopez N,
  Madueno E, De Pablos B, Wehland J, Kaerst U, Entian K, Hauf J,
  Rose M, Voss H;
WPI, 2002-010914/01.
XX Genomic sequence for Listeria monocytogenes, useful e.g. for treatment
PT and prevention of Listeria and related bacterial infections, and related
```

```
PT polypeptides.
XX
XX Claim 6; SEQ ID NO 39; 192pp; French.
PS
XX The present invention relates to the genome sequence of Listeria
CC monocytogenes EGD-e (see ABA03041). The genome sequence and fragments of
CC it are useful for selecting probes and primers for detecting genes in L.
CC monocytogenes and related organisms, and for studying genetic
CC polymorphisms and other genomes. The present sequence is a protein
CC encoded by the genome sequence of the present invention. Proteins
CC expressed from the genome sequence are useful for relating specific
CC antibodies, identification of L. monocytogenes and related organisms, and
CC for biosynthesis and biodegradation, especially biosynthesis of Vitamin
CC B12. The genome sequence and proteins encoded by it are also useful for
CC selecting compounds that regulate gene expression and cell replication
CC and modulate L. monocytogenes-related diseases. In addition, the genome
CC sequence and proteins encoded by it are useful in pharmaceutical and
CC vaccines compositions for the treatment or prevention of infections by L.
CC monocytogenes and related organisms. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
  ftp://wipo.int/pub/published_pcl_sequences
XX
SQ Sequence 940 AA;
Alignment Scores:
Pred. No.: 0.0885 Length: 940
Score: 108.00 Matches: 75
Percent Similarity: 37.03% Conservative: 52
Best Local Similarity: 21.87% Mismatches: 136
Query Match: 6.22% Indels: 80
DB: 5 Gaps: 16

US-10-617-962-2 (1-1008) x ABB47334 (1-940)
Qy 112 ACAGATCATATGAAATGAAATATGAGCT-----AAGAAATAATATTAAGCTTAC 162
Db 355 ThrValThrTyrLysAlaLysPheThrAlaThrSerLysGlyValProLeuSerTyr 374
Qy 163 GCTTGGCTGTAAGTGTATTCATATGATATCTAACTTCCGTAGAC----- 210
Db 375 SerIleAsnValSerGlnProIleAsnValSerGluGlnThrAspSerThrValSerVal 394
Qy 211 TATTTAAGAAATTAAGACAGCTGCGAGAAATTTATCAAGAAATATATGCTTAATCTTCA 270
Db 395 PheTyrGlnAspGluAsnGlyAsnGluLeuAlaProThrGluThrLeuSerGlyLysSer 414
Qy 271 TCTGCATATTAGTGGAAGAACGTCATCAAAATTTCTAAGAAATATGCAAAATGCTTTAT 330
Db 415 -----GlyGluAspTyrGlnThrThrGluLysThrIleAlaAsn----- 427
Qy 331 AAGATGAACGATTTGAAAGTCAATATCTCAAAACATTTGGAATGTTCTGAGCTT 390
Db 428 -----TyrGluLeuLysGluLe 433
Qy 391 GAAATTAACCATGAGTCTTATTCAGATGACGTAATATTAATATGACTATATTT--- 447
Db 434 GluGlyGlnAlaSerGlyGlnPheThrAspThrAspSerThrValThrTyrValTyrGlu 453
Qy 448 -----TTCTGTGACAGAGAAATTCACAGCTGAGAGAAATCAACATCA 489
Db 454 LysAlaAspGlyAlaProValThrValLysTyrValAlaAspGlyAsnAspLeuAla 473
Qy 490 AATGCCGCAAGATTTTAAATTAATTAATGTTTCTTATTAAGCTTTCGCTGTAATCTCA 549
Db 474 ThrSerAspThrLeuAsnGlyLysIleAspAlaProTyrGlnThrSerAlaLysSerLeu 493
Qy 550 CTGGAGAGAGGATT--TTTCAAAAACCTTTACAAATGATTA-----GAGCTTAA 600
Db 494 SerGlyTyrThrValLysThrThrProAsnAsnAlaThrGlyValPheThrAsnSerLys 513
Qy 601 TCATTAGGAATTAATATGAGAGAAATACTTTCTAATACCTTTTTCGACCAAGCGAG 660
```



```

Db      448 ProGluValIleThrGlnThrAsnProArgGlnAsnIleIleThrLysAsnAlaAsnHis 467
Qy      724 -----AGTTTAAAGAACTTAATAAACAATAATGACGATTTCTTAAT 771
Db      466 IleArgValAlaSerPheAsnValLeuAsnTrpAspAsnGlyAlaThrGlyPheProThr 487
Qy      772 ATGGAAGGGCGTCGCAAAACA-----AAGTAAATGTTCAATT 807
Db      468 GluArgGlyAlaAsnThrGlnAlaGluPheAspLysGlnHisIleLysValSerAla 507
Qy      808 ATAAAGAGGTA-----CAAAAGGTAAACGT 834
Db      508 LeuLysSerIleAspAlaAspValTrpGlyLeuMetGluIleAlaAsnAsnGlyTrpGly 527
Qy      835 CCACAGACGACGACGAAAAAGTATGTCAGCCAGCTGCG----- 873
Db      528 ProAsnSerAlaIleAlaHisLeuThrSerAlaLeuGlyProAspTrpLysTrpValIle 547
Qy      874 ---AGTAACTGGAAAAATTTGCCGAATTAATTAATTAATGTTGAGG----- 915
Db      548 ProGluAsnLeuAspArgLeuGlyAsnAspValIleAlaValAlaIleIleTrpAsnSer 567
Qy      916 -----CTAAGCCAAAAAGACAGGGTA 936
Db      568 LysArgValLysProLeuAsnLysAlaValValLeuAspLeuGlyAspLysAsnArgThr 587
Qy      937 ACCTTACTCA----- 948
Db      586 ThrLeuAlaGlnThrPheGlnAlaValArgGlyAsnLysIlePheThrValIleProAsn 607
Qy      949 -----AATGATACGTACAACTACA 966
Db      608 HisLeuLysSerLysGlyCysSerGlyValAspAlaSerSerSerAspAlaAspGln--A 627
Qy      967 ATGAGGTTTCATAGTCTGGAAC 989
Db      627 snAspGlyGlnGlyCysTrpAsn 634

RESULT 8
ADE83729
ID      ADE83729 standard; protein; 1270 AA.
XX
AC      ADE83729;
XX
DT      29-JAN-2004 (first entry)
XX
DE      Bacillus thuringiensis cry66 insecticidal protein #1.
XX
KM      Insecticidal protein; TK-E6; FERM P-17981; cry66; insecticide.
XX
OS      Bacillus thuringiensis.
XX
PN      JP2002355967-A.
XX
PD      26-NOV-2002.
XX
PF      16-MAY-2001; 2001JP-00146729.
XX
PR      16-MAY-2001; 2001JP-00146729.
XX
PA      (UYKI-) UNIV KINKI.
XX
DR      MPI; 2003-485827/46.
XX
DR      N-PSDB; ADE83728.
XX
PT      New Bacillus thuringiensis strain (Bacillus thuringiensis TK-E6 (FERM P-17981)) producing an insecticidal protein.
XX
PS      Claim 3; SEQ ID NO 2; 29pp; Japanese.
XX
CC      The invention comprises a microorganism that is capable of producing an insecticidal protein - Bacillus thuringiensis TK-E6 (FERM P-17981). The

```

```

CC      invention also comprises the amino acid and coding sequences of a number
CC      of insecticidal proteins (cry66 proteins). The bacterium and cry66
CC      proteins of the invention are useful as insecticides. The present amino
CC      acid sequence represents a Bacillus thuringiensis cry66 protein of the
CC      invention.
XX
SQ      Sequence 1270 AA;

Alignment Scores:
Pred. No.:      0.175      Length:      1270
Score:          105.50     Matches:      69
Percent Similarity: 34.19%   Conservative: 38
Best Local Similarity: 22.04% Mismatches: 119
Query Match:      6.08%     Indels:      87
DB:              7         Gaps:      14

US-10-617-962-2 (1-1008) x ADE83729 (1-1270)

Qy      13 TTAACACCTGATGATAGAACTGGATATCCACCCGTTGAAAAG----- 54
Db      544 LeuIleProAlaAsnLysIleGlyGlnIleProAlaValLysGlyAsnTrpLeuGlyGln 563
Qy      55 -----CAATAGACGAGATATGATGATATGATA 84
Db      564 SerGlySerSerValValArgThrSerGlyAsnThrGlyGlyAspValValArgLeuTrp 583
Qy      85 AACTTAAAGCAAAACAGATGAGGGTCATACAGCATCATATGGAATTAATGACCTAAG 144
Db      584 GluPheGlyAsnLeu-----GlyMetThrValArgPheSer--GluAsnArgSerTrp 600
Qy      145 AAAATTAATTAATTAAGCTTACCTTGGCTGTAAGTGATATCATATGTA----- 192
Db      601 IleIleArgLeuArgTrpAlaThrValAlaAspAspLeuAsnIleIleValArgValGln 620
Qy      193 -----TCTAACTCTCGATGACTATTTATTAAGATTAAGAG 228
Db      621 ArgArgGlyGluLeuGlnTrpGluSerGluLeuProLeuAsnGlnThrSerAsnAsnSer 640
Qy      229 ACTGCTGAGAGATTTATCAAGATATATGTTATCTTATCTTCATCTGCACTATTAAGTGAA 288
Db      641 ThrThrGlnTrpLysPheGluAspTrp----- 649
Qy      289 AATGCTGATCAAAATTTCTAAAGATATGCAAAATGTTTATTAAGAAATGACTGATTTT 348
Db      650 -----GluTrpGln-----GluVal 654
Qy      349 GAAGGCAATATATCCCTCAA-----AACATTGGGAATGTTCTGAGCTGAA 393
Db      655 GlyLysPheTrpProGlnValGlyGluGluTrpGluLeuTrpPheSerPro-----Val 672
Qy      394 AATAAACCATTTGAGTCTTATTCAGATGACGATTAATTAATTAAGCACTATATTTTCTCT 453
Db      673 GlyThrGluLeuSerSerHisMetAspIleAspLysIle-----GluPheIle 688
Qy      454 GTACAGCAAAATTCACCTGAGAGAA-----AATCAACATCAAAATGCCCAAGATTT 504
Db      689 ProMetGluTrpSerLeuGluGluTrpLeuAlaAsnGlnAspIleGluLysAlaArg--- 707
Qy      505 TTTTAATTAATTAATTTTATTTATTCCTTATCTGCTGTAATCTCACTGAGGAAGAGATT 564
Db      708 ---LysAlaValAsnAlaLeuPheThrGlyAspValLysAsnAlaLeuLysLeuAsnVal 726
Qy      565 TTTTCA-----AAAACTTTTACATGATGATTAAGAGCTTAATCTTAAGATTAAT 615
Db      727 ThrAspTrpAlaIleAspGlnAlaAlaAsnLeuValGluCysValSerGluGluPheTrp 746
Qy      616 ATTGAGAGAAAAAATTCTTAACCTTTCTTTGACCAACCGCAGAGATTAATCTGATGCG 675
Db      747 AlaGlnGluLysMetIleLeuLeuAspGlnValLysValAlaLysArgLeuSerGlnAla 766
Qy      676 AGAATAGCTTATTTGGCTGACCAACAGACGCGCTTAATG----- 717
Db      767 ArgAsnLeuLeuAsnTrpGlyAspPheGluSerProGluTrpSerArgGluAsnGlyTrp 786

```

QY 718 -----AGAGTGTGTTTAAAGACTTAAAGAAATACAAATCTAGAGATGATTT 765
 DB 781 LysThrSerArgHisValSerValArg---AlaAspAsnProValPheLysGlyArgTyr 805
 QY 766 TCTAATATGGAAGGGCTGCAAAACAAAGTATAGTTCA 804
 DB 806 LeuHisMetProGlyValThrSerProSerPheSerAsn 818
 RESULT 9
 ADK16573
 ID ADK16573 standard; protein; 2197 AA.
 AC ADK16573;
 DT 06-MAY-2004 (first entry)
 DE Nanoarchaeum equitans cancer-associated (CA) protein #262.
 XX
 XX cancer-associated gene; CA gene; cancer; carcinoma; lymphoma; leukaemia.
 OS Nanoarchaeum equitans.
 XX
 XX WO2003093434-A2.
 XX
 XX 13-NOV-2003.
 XX
 PF 01-MAY-2003; 2003WO-US013699.
 XX
 PR 01-MAY-2002; 2002US-0377447P.
 XX
 PA (DIVE-) DIVERSA CORP.
 PI Stetter KO, Waters E, Kretz K, Podar M, Richardson T,
 PI Noordewier M;
 XX
 DR WPI: 2004-053041/05.
 DR N-PSDB; ADK16572.
 XX
 PT New recombinant cancer-associated genes, such as KCNU9, useful for
 PT diagnosing or treating carcinoma, e.g. breast, colon, rectal, pancreatic,
 PT cervical, or skin cancers, lymphomas, or leukemia.
 XX
 PS Claim 64; SEQ ID NO 525; 251bp; English.
 XX
 CC The invention comprises then amino acid and coding sequences of cancer-
 CC associated (CA) genes isolated from Nanoarchaeum equitans. The invention
 CC also comprises the Nanoarchaeum equitans genome. The DNA and protein
 CC sequences of the invention are useful for diagnosing and treating cancer
 CC (e.g. carcinoma, lymphoma, or leukaemia). The present amino acid sequence
 CC represents a Nanoarchaeum equitans CA protein of the invention.
 XX
 SQ Sequence 2197 AA;
 Alignment Scores:
 Pred. No.: 0.262 Length: 2197
 Score: 104.50 Matches: 72
 Percent Similarity: 38.07% Conservative: 54
 Best Local Similarity: 21.75% Mismatches: 154
 Query Match: 6.02% Indels: 51
 DB: 8 Gaps: 13
 US-10-617-962-2 (1-1008) x ADK16573 (1-2197)
 QY 73 GTAAGTATACCTTAACCTTAAGCAACAGATGAGGCTCATACAGATCATATGGAATTGA 132
 DB 670 ValLeuLeuSerAsnIleArgAsnAsnSerTyrAlaThrLeuSerAsnTyrLeuIleHis 689
 QY 133 TATCGAGCTTAAGAAATATATATAGCTTACGCTTGGCTGTAAGTGTATCATATGTA 192
 DB 690 TyrAlaAlaTyrIleValIleAsnAspIleValAsnGluIleIleAsnGluLeuAsnIle 709
 QY 193 TCTAAACTTCGTGATGACTAT-----TATTAAGATAAAGAGACTGTAAGAAATTAT 246

DB 710 -----AapAspPheValGlnTyrAsnLeuLys-----LysGluLeuIle 722
 QY 247 CAAGATATATAGCTTAACCTTTCATCTGCACATATAGCTGAAGAAATGATCAATTTCT 306
 DB 723 LysAsnTyrPValSerAsnSerIleSerHisCysLeuGlyGlu----- 736
 QY 307 AAAGATATGCAAAATGTTTATTAAGAATGACATGATTTTGAAGTCAATATCTGCA 366
 DB 737 ValAspIleLeuMet---PheLeuThrAlaAsnIleAspLeuGluArgLeuTyrGln 755
 QY 367 AACATTTGGAATGTTCTGACCTTGA--ATAAACCATTTAGTCTTATTCAGATGAC 423
 DB 756 GluLeuGlnAsnTyrIleAlaLeuSerValLeuThrProLeuSerIleTyrSerSerAsn 775
 QY 424 GATTAATTTATGACATATATTTTCTCTGACAGCAAAATCCATGAGCAAAATCA 483
 DB 776 Glu-----ValLysTyrIleArgGluMetValLysSerGlnLeuGluMetGlyLeu 792
 QY 484 CAATCA-----ATGCGCGCAAGATTTTAAATTAAT 516
 DB 793 SerSerLeuAspAsnLysLeuGluProIleIleAsnThrLeuThrAsnSerLeuTyr 812
 QY 517 GATTTCTTATTTACCTTATCTGCTGTAACTTCACTGAGAGAGATTTTCAAAAAC 576
 DB 813 AsnLeuAlaTyrSerLeuSerGlnValThrSerSerLeuThrAsnTyrPheSer---Ser 831
 QY 577 TTTTCAATGATTTAGAGCTTAATCATTTAGAGATTTATTTAGAGAAAAAACTT--- 633
 DB 832 ValValSerSerIleAsnSerLeuValLeuProSerThrIleGlyGluIleLysLeuIle 851
 QY 634 -----TCTAAACCTTTCTTGACACCGACGAGATTCGTATGCGAGAATA 681
 DB 852 AspTyrLysThrLeuLysAlaIleAlaLeuIleProSerSerValGluGlnAlaIleIle 871
 QY 682 GGTATTTTGCTGACCAACAGACGCGCTTAATAGAGAGTGTATTAAGAACTTAA 741
 DB 872 AsnArgLeu-----ThrAsnLysValSerTyrAsnIleLeuLeuGluCysIleAsp 888
 QY 742 AATAACAAATCTAGAAATGATTTCTTAATATGCAAGGGCTGCAAAACAAAGTAT 801
 DB 889 ThrLysLeuSerLysAlaAlaAsnAlaGlnIleAlaLeuAlaLysLeuGlyTyrLeu 908
 QY 802 TCATTATTAAGAGATCAAAAGGTAAAGCTGACAGACAGACGCAAAAGTATGAT 861
 DB 909 SerIleIleGluGlyIleAspProAspSerTyrProAspLysLysAlaGln----- 925
 QY 862 ACAGCAGTGGCAGTAACTGGAATAATGCCGAATAATTTATATAGTGTGAGCTTAAC 921
 DB 926 -----LeuGluGlnLeuIleSerLysValLysGluValLysLysIle 939
 QY 922 CAAAAAGCAGGGTAACTTTACTCAAAATGATCTGACATACATACATGACGTTCTAT 981
 DB 940 IleGluGlnGlnGlnSerThrLeuLysLysAspLeuAspAsnThrIleAsnLeuTyrLys 959
 QY 982 GTT-----GGAACCTATTAATAAAT 1002
 DB 960 GluPheIleLysGlySerThrGluTyrIleAsn 970
 RESULT 10
 ABR63439
 ID ABR63439 standard; protein; 583 AA.
 XX
 AC ABR63439;
 DT 08-SEP-2003 (first entry)
 DE S suis aro gene cluster encoded protein #1.
 XX
 XX Avirulent vaccine; antibiotic; aro gene cluster; immunisation;
 KW antibacterial; gene therapy.
 XX
 OS Streptococcus suis.

```

XX  MO2003046183-A2.
XX  PD 05-JUN-2003.
XX  PF 25-NOV-2002; 2002WO-CA001796.
XX  PR 23-NOV-2001; 2001US-0332012P.
XX  (UWMO-) UNIV MONTREAL.
XX  Gottschalk M, Harel J, D'amours B, Kobish M;
XX  WPI, 2003-505203/47.
XX  N-PSDB; ACC59/57.
XX  DR  New isolated or recombinant nucleic acid encoding an aro gene cluster of
XX  PT Streptococcus suis, useful for controlling or eradicating streptococcal
XX  PT diseases.
XX  Claim 3; Fig 7; 49pp; English.
XX  PS
XX  CC The present invention provides the protein and coding sequences of
XX  CC sequences contained within the Streptococcus suis aro gene cluster. The
XX  CC sequences are useful for controlling or eradicating a streptococcal
XX  CC disease in a population by administration to subjects of the population.
XX  CC The strain is useful in raising polyclonal antibodies against the mutant
XX  CC in a host. Inhibitors and/or analogues of the aromatic biosynthesis
XX  CC pathway are useful in preventing the expression of capsular
XX  CC polysaccharides by pathogenic Streptococci. The present sequence is a
XX  CC protein encoded by a gene from the S. suis aro cluster
XX  SQ Sequence 583 AA;

Alignment Scores:
Pred. No.: 0.197 Length: 583
Score: 104.00 Matches: 85
Percent Similarity: 40.73% Conservative: 60
Best Local Similarity: 23.88% Mismatches: 128
Query Match: 5.99% Indels: 83
DB: Gaps: 21

US-10-617-962-2 (1-1008) x ABR63439 (1-583)
QY 49 GAAGACCAATATAGCAGAGATATAGTACTATCTA-----AACTTAAAGCAACA 99
DB 152 GlnTyrGlnTyrAsnTyrSerIleIleValIleGlySerGlnLeuTyrAsnAsn 171
QY 100 GATGAGGTCATACACATCATATGGAATT-----GAATATGAGCTAAGAAATA 150
DB 172 AspIleAsxPheSerTyrAsnAsnAsnValHisIlePheGlnLeuAsnAsnAsn 191
QY 151 ATATTAGCTTACGCTTGGCTGTAAAGTATGATATCATATGATCTTAACTTCTGATGAC 210
DB 192 ThrIleGlyAsnIleIleGlnThrSerIleAsnIleSerAsp-----AsnLeuGlnAspThr 209
QY 211 TATTATTAAC--AATAAAGAAGCTGCTGAGAGATTTATCAAGATATATATGTTCTT 267
DB 210 LeuTyrProLeuThrTyrGlyAsnLeuGlnTyrIleTyr----- 222
QY 268 TCATTCGACACTATTAAGTAAATGATGATCAATTTCTAAA-----GATATGCA 318
DB 223 AsnPheIleLeuLeuSerGlnAsnAsnSerLysIleGlnTyrLeuAsnAsnAsn 242
QY 319 AAT-----GGTTTTTATAGAAATGAAGCTGGAATTTTGAAGCTCAATATCTCTCA 366
DB 243 AsnThrIleGlnIleAsnIleAsnIleGlnIleAsnIleGlnIleAsnIleTyrGlnLeu 262
QY 367 AACATTGGAATGCTTCTGAGCTTGAATAATTAACATGAGTCTTATTCAGATGACGAT 426
DB 263 SerVal-----LeuProGlnTyrPheAsnPro-----TyrAspAlaGln 275
QY 427 AATATTATGACACTATATTTTCTCTGTACAGAA----- 462

```

```

DB 276 GlnTyrPheAsnLeuTyrThrPheGlnIleGlnTyrAlaIleAspIleLeuTyrLysLeu 295
QY 463 -----ATTCCACTGGAGGAAATACACATCAATCAAAATCCGCAAGATTTTAAATTA 513
DB 296 LeuValLeuIleProValLys-----AanArgThrAsnPheTyrSerLeu 310
QY 514 ATT-----GATTTCTTATTAATCTTATCTGCTGCTTAATCTTCACTGGAGAG 558
DB 311 IleLeuIleMetLysAspLysLeuLeuLeuThrValIleGlnTyr-----GlyGln 326
QY 559 AGGATTTTTCAAAAAAGCTTTTACAAATGATAGAGGCTAAATCATATGAGAATATATAT 618
DB 327 LysGlnIleTyrArgTyrTyrHisTyrLeuTyrGlyLysGlnProPheAsnTyr--- 345
QY 619 GAGAGAAAAAAGCTTTCTTAACCTTTCT-----TTCAGCACCGCAGAGATTAATCTGATGAC 675
DB 346 ---ArgLeuLysIleGlnIleLeuSerArgLeuLysGlnSerGlnGlnTyrLeuLeu 364
QY 676 AGAATAGCTTATTTGGCTGGACCAACAGAGCGCTTAATGAGAGTGAATTTTAAAGA 735
DB 364 rLeuTyrLeuLeuTyrTyrGlnTyr-----LeuLeuAsnGlnAsnGlnIleLeuGln 382
QY 736 CTTAAATAATTAACAAATCTAGAAATGATTTCTAATATGAAAGGGGCTGCAAAACAAAG 795
DB 382 rGlnLysLysSerLeuTyrLysLeuGlnPheLeuThr-----GlnGlnThrLys 399
QY 796 TTTAGTTCATTTTATTAAGAAGGTATCAAAAGGTAAAGCTCCACAGACAGACGAAAGT 855
DB 399 rIle-PhePhe-----LysCysIleLysT 407
QY 856 ATTGGTACAGCCAGTGGCAGTAACTG-----GAAAAATTCGCCAATATTTATAT 906
DB 407 TLTyrTyrThr-AsnGlyThrAsnSerLeuTyrAsnGlnIleLeuProGlnLysSerLeu 426
QY 907 AGTGTAGAGCTTAAGCAAAAGACAGAGGTAACTTTACTCAAAATGATCTGACAATACA 966
DB 427 LeuSerSerIleLeuLeuLysAsnAsnAsnIleAsnPhe---GlnGlyAsnThrSerGlnAsp 445
QY 967 ATGACGGTTCAATATGTTT---GGAACGTATTTAAAAATATA 1005
DB 446 PheSerPheHisSerLeuTyrAsnAspThrLeuTyrGlnAsnIle 459

RESULT 11
AAM54145
ID AAM54145 standard; protein; 1639 AA.
AC AAM54145;
XX
XX 25-MAR-2003 (revised)
XX 23-SEP-1998 (first entry)
XX
XX P. falciparum synthetic gp190 protein.
XX
XX gp190; malaria; MSP-1; merozoite surface protein; stability; vaccine;
XX monoclonal antibody; passive immunisation; parasite.
XX
XX Plasmodium falciparum.
XX Synthetic.
XX
XX OS
XX PN WO9814583-A2.
XX
XX PD 09-APR-1998.
XX
XX PF 02-OCT-1997; 97WO-BP005441.
XX
XX PR 02-OCT-1996; 96DE-01040817.
XX
XX PA (BUA/) BUARD H.
XX
XX Bujard H, Tolle R, Pan W;
XX
XX WPI, 1998-240086/21.
XX

```

DR N-PSDB; AAV21451, AAV35363.

XX Recombinant production of complete gp190/MSP-1 Plasmodium surface protein
PT - useful in anti-malaria vaccines, also stabilising genes by reducing
PT their AT content.

XX Example 1; Fig 3c; 48bp; German.

XX This sequence represents a modified plasmidium falciparum gp190/MSP-1
CC (merozoite surface) protein. The gene encoding this protein has been
CC stabilised by reducing the AT content of the nucleotide sequence. Such a
CC protein is useful in vaccines against malaria or for producing monoclonal
CC antibodies (for passive immunisation). The complete gp190 protein can now
CC be produced outside the parasite and has, at least over extended regions,
CC the native pattern of folding. Larger amounts of the protein can be
CC produced recombinantly than would be possible using the parasites as
CC source. (Updated on 25-MAR-2003 to correct PR field.)

XX SQ Sequence 1639 AA;

Alignment Scores:

Pred. No.:	0.487	Length:	1639
Score:	101.50	Matches:	82
Percent Similarity:	34.70%	Conservative:	53
Best Local Similarity:	21.08%	Mismatches:	137
Query Match:	5.85%	Indels:	117
DB:	2	Gaps:	19

US-10-617-962-2 (1-1008) x AAW54145 (1-1639)

```

QY 121 TATGGAATGATATCGAGCT-----AGAAAATATATTAGCTTAC 162
DB      ::::: |||||
DB 205 Pheanleuyllelghlaenleuaspvalleuylleuvalhegllyr 224
QY 163 GCTTGGCTGTAAGTGATTT--CATATGATCTTAACTCTGATGACTTATTAAG 219
DB      ::::: |||||
DB 225 ArglysproleuaspnlellyspenValglyMetcluhapryrillelyls 244
QY 220 AATAAGACCTGCTGAG-----
DB      |||||
DB 245 Aenlyllyethrileghuanlleasnleuileugluserlylserthrlleap 264
QY 238 -----AGATTATCAA---GATATATG 258
DB      ::::: |||||
DB 265 LysaenlyspasnAlathrlysegluglulyslylserlyglhlaelintyrp 284
QY 259 TCTAATCTTTCATCTGCACTATTAGTGTAATAATGATCAATTTCT-----AAAGAT 312
DB      ::::: |||||
DB 285 LeuserlleTyAsnlysegluleugluahlaenleuilesevalleugluys 304
QY 313 ATGGCAATGCTTTTATAGATGA-----CTGGAT----- 345
DB      ::::: |||||
DB 305 Argileasprthleuyllyspasnleuileugluleuaspysilleasnlu 324
QY 346 TTTGAAGGTCATATCTCAAAACATTGGATGTTCCGAGCTT-----GAAAT 396
DB      ::::: |||||
DB 325 IlelyspasnproforlaahnsenrglyasnThrproasnThrleuaspysasn 344
QY 397 AAACCATTTGAGTCTTATTCAGATGACATTAATATTATGACTATATTTT---TTCTCT 453
DB      ::::: |||||
DB 345 Lyslyslleugluhileuglulysglulileuilehlythrilleyspheaen 364
QY 454 GTACGGAATTT-----CCACTG----- 471
DB      ::::: |||||
DB 365 IleasprleuethrleasprleuugluelyrTyrlleuarglulysasnly 384
QY 472 -----GAGAAATCAACATCA 489
DB      ::::: |||||
DB 365 AsnilleasprleuSerAlalyValgluthrlysgluserThglulproasnleuTyrrpo 404
QY 490 AATGCCGCAAGATT---TTTAAATTAATGATTTCTTATTTACCTTATCTGCTGTA 546
DB      ::::: |||||
DB 405 AenglyValThrTyrrProleuSerTyraenAsprlleasnAlaleuasnleu 424

```

```

QY 547 TCACGTGGAAGAGGATTTT-----TCAAAAACTTTTAC 582
DB      |||||
DB 445 SerpneghlyaspneulleasnproheasprlyrthrlysgluProserlyspasnlelyr 444
QY 583 AATGATTAAGAGCTTAATTCATTAGCAATATATATTGAGAGAAA---AACTTTCTAA 639
DB      |||||
DB 445 Thrleasngluarglyllyspheilleasnleuileuglulyleytleleuglulys 464
QY 640 CCTTTCTTTCGACACCGCAGAGATTACCTGACAGAAATAGTTATTGGCTGACCA 699
DB      ::::: |||||
DB 465 LyslyslleugluseraspyspserlyrGlueasprlysserleuasnaspile 484
QY 700 ACAGAGGCGCTTAATGAGAGTGATTTTAAAGACTTAAATAACAAATCTAGAT 759
DB      ::::: |||||
DB 485 Thrlys-----gluTyglulysleuasnleuileuileuTyrrasprlyspheasn 502
QY 760 GGA-----TTTTTAATATGGAAGGCGCTGCAAAACAAAGATATGTTATTAATA 813
DB      ::::: |||||
DB 503 AsnilleasprleuThrAsnPhelglulysMetetgllylsargTyrrserlyrlyValglu 522
QY 814 GAGGTACAAAAGGTTAACCTCCACAGACAGACGAAAGATTTGTACAGCCAGTGC 873
DB      ::::: |||||
QY 523 LysleuThrhlshlspasnThrPheAlaSerTyrgluasn-----Serlys 537
DB 874 AGTAACTGGAAGAAATTCGCCGAATATTTA-----TATAGTGTAGAGCTA 918
DB      |||||
DB 538 hlsasnleuuglulysleuThrlyslaleuylsTyrrMetgluasprlyserleuarg--- 556
QY 919 AGCCAAAAAGACAGGTTAACTTACTCAAAATGATGACAAATACATGACGGTTCAT 978
DB      ::::: |||||
DB 557 -----AsnilleValglu 561
QY 979 AGTGTGCACTCATTTAATAATATA 1005
DB      ::::: |||||
DB 562 LysgluleuylsTyrrTyrllyspasnleu 570

```

RESULT 12

AAE29345

ID AAE29345 standard; protein, 1639 AA.

AC AAE29345;

XX 27-JAN-2003 (first entry)

XX DT

XX 27

XX DE Plasmodium falciparum merozoite surface protein-1 (MSP-1).

XX KW Band 3 polypeptide; malarial infection; drug resistance; vaccine;

XX KW prozoocicide; gene therapy.

XX OS Plasmodium falciparum.

XX OS

XX PN WO200270542-A2.

XX PD 12-SEP-2002.

XX PF 01-MAR-2002; 2002WO-US006415.

XX PR 02-MAR-2001; 2001US-0272930P.

XX PR (SELI-) ST ELIZABETH'S MEDICAL CENT INC.

XX PA

XX PI Chienti AH, Oh SS, Liu D, Goel V;

XX PI

XX DR WPI; 2002-759814/82.

XX DR N-PSDB; AAD46980.

XX PT New isolated Band 3 polypeptide which selectively binds to merozoite

XX PT surface protein-1, useful for the prevention and treatment of malarial

XX PT infection.

XX PS Disclosure; Page 113-118; 163pp; English.

CC which is attached thereto from cells. Such protein complexes may comprise
CC up to 30 distinct proteins. Protein complexes of the invention are useful
CC for diagnosing a disease or disorder, or as a target for an active agent
CC of a pharmaceutical, preferably a drug target in the treatment or
CC prevention of a disease or disorder. Note: The sequence data for this
CC patent is not represented in the printed specification, but is based on
CC sequence information supplied by the European Patent Office. The complete
CC document is available on CD-ROM

XX
SQ Sequence 566 AA:

Alignment Scores:
Pred. No.: 0.396 Length: 566
Score: 101.00 Matches: 46
Percent Similarity: 38.01% Conservative: 38
Best Local Similarity: 20.81% Mismatches: 87
Query Match: 5.82% Indels: 50
Gaps: 8

US-10-617-962-2 (1-1008) x ABR53129 (1-566)

QY 442 TATTTTCTCTGTACAGAAATTCCTGAGGAAATCAACATCAATCCGCAAGA 501
DB 27 PhePheH1sAlaLylsLysAspProValaEnglnAspLysAlaAsnAsnLaseGrgln 46
QY 502 -----TTTTTT 507
DB 47 IleHrProThrValProHisSerHisProSerAspMetValIleProAspHisLeuAla 66
QY 508 AAATTAATGATTTCTTATTTATCTTATCTGCTGTAATCTTCACTGGAAGAGATTTT 567
DB 67 GluLeuIleProGluLeuTyrSerPheGlnGlnLeuValAspSerGluLysArgLeuAsp 86
QY 568 TCAAAAACCTTTTACATGATGATTAGAGCTAAATCATTAAGAAATTATTAAGAGAAA 627
DB 87 HisPheIleHisLeuAlaGlnAsnLeuHisMetLysArgMetValAlaGlnTrpGluArgSer 106
QY 688 AAACCTTTTAAACCTTTCTTGTGACCCGACAGATTAATCGAGGAGCAATA----- 681
DB 107 LysLysSerGlnGlnPheLeuTyrProHisLeuAsnPheProAsnValAspPheLeuArg 126
QY 682 GGTATTTTGGCTGACCAACAGAGAGCGCTTAATGAGAGATGTTTAAAGAACTTAA 741
DB 127 IlePheIleSerAsnValSerGlnAsnGlnProTrpGlnMet-----AspThrAsn 143
QY 742 AATAACAATCTAGAAATGATTTCTTAAT-----ATGAGAGG--- 780
DB 144 AsnGluAlaAspLeuMetAlaLeuGlnAsnAlaThrTrpThrMetArgIleGluArg 163
QY 781 -----GCTGCAAAAACAAAGATATGATTTTAAATA 813
DB 164 LeuLeuAspAsnValGlnAlaAsnAspProAlaArgLysGlnLysPheSerPheIleGlu 183
QY 814 GAGGTA-----CAAAAGGCTACGCTCCACAGACA----- 843
DB 184 SerIleValValaAspPheLysAsnLysGlnAsnAspAsnValProSerThrLysPheAsn 203
QY 844 GCAAGCAAAAATGATTGGTACAGCCAGTGGCACTGAGAAAATTCGCAATATATTA 903
DB 204 AlaAlaProGluGlnAsnAlaThrGlnGlyProSerAspLysLysLeuAsnLeuAsn 223
QY 904 ---TATAGTGTAGGCTAAGCAAAAAGCAGGCTAATCTTTATCAAAATGATCTGC 960
DB 224 ProLeuGlnPheSerLeuProAsnGlnLysAsnSerThrThrThrAsnThrAspGlnAsn 243
QY 961 AAT 963
DB 244 Asn 244
RESULT 14
ADK62560
ID ADK62560 standard; protein; 566 AA.
XX

AC ADK62560;
XX 06-MAY-2004 (first entry)
DT
XX Disease treating protein complex-derived protein #406.
DB
XX protein complex; drug target; diagnosis.
XX
XX Unidentified.
XX
XX EP1338608-A2.
XX
XX 27-AUG-2003.
XX
XX 20-DEC-2002; 2002EP-00102902.
XX
XX 20-DEC-2001; 2001EP-00130253.
XX
XX (CELL-) CELLZOME AG.
XX
XX Bauer A, Gavin A, Superfi-Furga G, Kuester B, Schultz J;
PI Marzoch M, Grandi P, Krause R, Krause U, Merino A, Bauch A;
PI Michon A, Leutwein C, Rick J;
XX
XX WPI, 2003-638460/61.
XX
XX N-PSDB; ADK62561.
XX
XX
XX New proteins and protein complexes from eukaryotes, useful as targets in
PT drug screening, or in diagnosing or screening for the presence of a
PT disease or disorder, or a predisposition for developing a disease or
PT disorder in a subject.
XX
XX
XX Disclosure; SEQ ID NO 811, 13pp; English.
XX
XX The invention relates to novel protein complexes comprising a first and a
XX second protein, or its derivative, fragment, homologue or variant. The
XX proteins are selected from given protein complexes, which are not defined
XX in the specification. The variants are encoded by nucleic acids that
XX hybridize to the nucleic acids encoding the proteins under low stringency
XX conditions. The protein complexes are useful as targets for an active
XX agent of a pharmaceutical. These protein complexes are particularly
XX useful as drug targets for the treatment or preventing of a disease or
XX disorder. The complexes and methods above are useful in diagnosing or
XX screening for the presence of a disease or disorder or a predisposition
XX for developing a disease or disorder in a subject. These are also useful
XX in screening for a drug for treatment or prevention of a disease or
XX disorder. The molecule that modulates the amount, activity or protein
XX components of the complex is useful for the manufacture of a medicament
XX for the treatment or prevention of a disease or disorder. This sequence
XX corresponds to a protein of the invention. (Note: the sequence data for
XX this patent did not form part of the printed specification but was
XX obtained from the EPO in electronic format).

SQ Sequence 566 AA:

Alignment Scores:
Pred. No.: 0.396 Length: 566
Score: 101.00 Matches: 46
Percent Similarity: 38.01% Conservative: 38
Best Local Similarity: 20.81% Mismatches: 87
Query Match: 5.82% Indels: 50
Gaps: 8

US-10-617-962-2 (1-1008) x ADK62560 (1-566)

QY 442 TATTTTCTCTGTACAGAAATTCCTGAGGAAATCAACATCAATCCGCAAGA 501
DB 27 PhePheH1sAlaLylsLysAspProValaEnglnAspLysAlaAsnAsnLaseGrgln 46
QY 502 -----TTTTTT 507
DB 47 IleHrProThrValProHisSerHisProSerAspMetValIleProAspHisLeuAla 66

```

QY 508 AATATATGATTTCTTATTACCTTATGCTGTAATCTTACCTGGGAAGAGATTTT 567
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 67 GlutelleProgluLeuTySerPheGlnGlnLeuValAspSerGluLysArgLeuAsp 86
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 568 TCAAAAACCTTTACATGATGATAGAGGCTTAATCATCTAGAGATTATATTAGAGAAA 627
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 87 HisPheIleHisLeuArgAsnLeuHisMetLysValMetValAlaGlnTyrGluLysSer 106
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 628 AAACCTTTTAAACCTTTCTTTGAGACCGACGAGATTCTGATGGCAGAAATA----- 681
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 107 LysLeuSerGlnGlnLysPheLeuTyrProHisLeuAsnPheProAsnValLysPheLeuArg 126
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 682 GGTATTATTTGCTGACCAACAGAGCGCTTAATGAGAGTGAAGTTTAAAGAACTTAA 741
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 127 IlePheIleSerAsnValSerGluAsnGlnProTyrGlnMet-----AspThrAsn 143
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 742 AATAACAATCTAGAAATGATTTCTAT-----ATGGAAGGG----- 780
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 144 AsnGluAlaAspLeuMetAlaLeuGluAsnAlaThrTyrThrMetArgIleGluLysArg 163
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 781 -----GCTGCAAAACAAAGATATAGTTTCAATTATTA 813
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 164 LeuLeuAspAsnValGlnAlaAsnAspProAlaArgGluLysPheSerSerPheIleGlu 183
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 814 GAGGTA-----CAAAAGGTAACGCTCCACAGCA----- 843
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 184 SerIleValValAspPheLysAsnLysGluAsnAspAsnValProSerThrLysPheAsn 203
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 844 GCAGCGAAAGATTTGTTACGACGACGAGTGAACCTGGAATAATTCGCAATATTA 903
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 204 AlaAlaProGluGluAsnAlaThrGluGlyProSerAspLysLysLeuAsnLeuAsnLeu 223
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 904 ---TATAGTGTAGGCTTAAGCCAAAAGAAGAGGTAACTTTACTCAAAATGATACTGAC 960
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 224 ProLeuGlnPheSerLeuProAsnGlnLysAsnSerThrThrAsnThrAspGlnAsn 243
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 961 AAT 963
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 244 Asn 244
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 15
ADP25426
ID ADP25426 standard; protein; 2404 AA.
XX
XX ADP25426;
XX
AC 09-SEP-2004 (first entry)
XX
DT 09-SEP-2004 (first entry)
XX
DE Plasmodium falciparum antigen amino acid sequence SEQ ID NO:3.
XX
KW Plasmodium falciparum; malaria parasite; antigen; malaria; immunogenic;
KW immune response; cytotoxic; anti-HIV; virucide; hepatotropic;
KW antibacterial; vaccine; cancer; infectious disease; AIDS; hepatitis;
KW bacterial infection.
XX
OS Plasmodium falciparum.
XX
PN WO2004053086-A2.
XX
PD 24-JUN-2004.
XX
PF 08-DEC-2003; 2003WO-US038966.
XX
PR 06-DEC-2002; 2002US-0431494P.
XX
PA (EPIM-) EPIMUNE INC.
XX (USNA ) US SEC OF NAVY.
XX
PI Sette A, Doolan DL, Carucci DJ, Sidney J, Southwood S;
XX
XX WPI; 2004-468856/44.
XX
XX New isolated and/or purified Plasmodium falciparum polynucleotide

```

```

PT sequences, useful in inducing an immune response for preventing and/or
PT treating cancer and infectious diseases, such as AIDS, hepatitis, and
PT bacterial infections.
XX
XX Claim 22, SEQ ID NO 3; 253pp; English.
XX
CC The present invention describes an isolated and/or purified Plasmodium
CC falciparum (malaria parasite) antigen polynucleotide sequence, encoding
CC an immunogenic peptide. Also described: (1) a primer or detection probe
CC for hybridisation with a target sequence or the amplicon generated from a
CC target sequence comprising a sequence of at least 8-30, 35, 40, 45, 50,
CC 55, 60, 65, 70, 75, 80, 85, 90, 95 or 100 consecutive nucleotides of any
CC of the polynucleotide sequences as described above; (2) a DNA chip
CC comprising any of the polynucleotide sequences described above; (3) a
CC vector comprising a promoter operably linked to any of the nucleic acid
CC sequences described above; (4) a host cell transformed by the vector of
CC (3) or the polynucleotide described above; (5) a composition comprising a
CC carrier and the polynucleotide described above; (6) a method of inducing
CC an immune response in an individual comprising the administration of the
CC composition of (5) to induce an immune response; (7) an isolated
CC polypeptide comprising any of the amino acid sequences as encoded by the
CC polynucleotide described above; (8) a composition comprising a carrier
CC and the polypeptide of (7); (9) a method of detecting P. falciparum in
CC biological samples, comprising contacting a biological sample with the
CC isolated polynucleotide and detecting the hybridisation of the isolated
CC polynucleotides with nucleic acids contained in the sample; (10) a method
CC for eliciting an immune response in an individual, comprising the
CC administration of a composition comprising the polypeptides of (7) to an
CC individual to induce an immune response in the individual; (11) an
CC antibody that specifically binds to the P. falciparum polypeptide of (7);
CC and (12) detecting P. falciparum antigens, comprising contacting a sample
CC from a subject with the polypeptide of (7) and detecting the presence of
CC an antigen-antibody complex or detecting the stimulation of T-cells in
CC the sample. The P. falciparum antigens and immunogenic peptides have
CC cytosolic, anti-HIV, virucide, hepatotropic and antibacterial
CC activities, and can be used in vaccines. The methods and compositions of
CC the present invention are useful for inducing an immune response for the
CC prevention and/or treatment of cancer and infectious diseases, such as
CC AIDS, hepatitis, and bacterial infections. The present sequence
CC represents a P. falciparum antigen amino acid sequence, which is used in
CC the exemplification of the present invention.
XX
XX SEQ Sequence 2404 AA;
XX
XX
XX Alignment Scores:
XX Pred. No.: 0.779 Length: 2404
XX Score: 100.00 Matches: 49
XX Percent Similarity: 41.95% Conservative: 37
XX Best Local Similarity: 23.90% Mismatches: 54
XX Query Match: 5.76% Indels: 65
XX DB: 8 Gaps: 10
XX
XX US-10-617-962-2 (1-1008) x ADP25426 (1-2404)
QY 142 AAGAAATATATGATTTAGCTTACGCTTGGCTGTAATCTTATCT----- 186
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 326 LysAsnThrIleAsnAlaPheSerIleAsnTyrThrGlyValAsnPheGluAlaMetLys 345
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 187 -----AATGATCTAACTTCCTGATGAC---TATTAAGATTAAGAGACT 231
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 346 GluLeuAsnAspLysAlaSerLeuLeuPheAspAsnValTyrTyrGluLysLysGluAsn 365
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 232 GCTGAGAGATTTATCAAGATATATATGCTTATCTTATCTGACATTAAGTGGAAT 291
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 366 SerAsnArg-----GluGluIle 371
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 292 GGTGATCAAAATTTCTAA-----GATATGGCAAAATGCTTTT 327
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 372 AsnAspLysValSerLysGlnGlyCysAsnLeuAsnAspSerAspSerSerAsnValLeu 391
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 328 TAT-----AAGATGAAGTGAATTTTGAAGTCAATATCCCAAAACATTTGGAAT 378
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 392 TyrIleAsnIleGlnAsnIleLysAspTyrAspIleLeuTyrLysGluAspAsnLysAsn 411
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```
OY      GTTCTGAGCTGGAAATAAACCATTGATGCTTATCAGAT----- 420
Db      379   TTTAAAGTGAAGTAAAGTAAACCAATGATGCTTATCAGAT----- 420
OY      412   TTTAAAGTGAAGTAAAGTAAACCAATGATGCTTATCAGAT----- 420
Db      412   TTTAAAGTGAAGTAAAGTAAACCAATGATGCTTATCAGAT----- 420
OY      421   ---GACGATTAATTTATTAGACATATTTTTCTGTGCAGGAAATTCACATGAGGAA 477
Db      432   ValAspLeuLysAsnMetAlaLeuHisIlePhe----- 442
OY      478   AATCAACAATCAAAATGCCCGCAAGATTTTAAAATTGATTTCTTATTTACCTTATCT 537
Db      443   -----PheTyrllysIleIleasp-----GluThrGln 451
OY      538   GCTGTAACTTCACTGGGAAAGAAGATTTTTCAAAAACTTTTACAATGATGATTAGAAGCT 597
Db      452   HisValValHisMetAsnLysLysGluTy--LysTyPheHisLeuValMetLysIle 470
OY      558   AAATCATTTAGAGATTTATTTAGAGAGAAAA---AACTTTCTTAACCTTTCTTGACCA 654
Db      471   LeuPheLeuHisAsnTyLr-LeuLeuGlnLysMetAsnMetLeuAsnLeucylsIleAspAs 490
OY      655   CCGCAGAGATTAC 667
Db      490   nLeuAsnGlnPhe 494
```

THIS PAGE BLANK (USPTO)

Result No.	Score	Query Match	Length	DB	ID	Description
1	109.5	6.3	414	2	H70219	hypothetical protee
2	108.5	6.3	1640	2	A24544	probable major sur
3	108	6.2	940	2	AD13174	intermalin protein
4	108	6.2	1127	2	T28317	ORF MSV156 hypothe
5	106	6.1	1785	2	A45546	major mekrotoxin su
6	104.5	6.0	755	2	T41912	structural phospho
7	104	6.0	1428	2	C85079	hypothetical prote
8	104	6.0	2708	2	T09079	probable chloroquin
9	103	5.9	579	2	H97194	uncharacterized AB
10	103	5.9	853	2	S60178	gas polyprotein ho
11	103	5.9	950	2	T09076	hypothetical prote
12	102.5	5.9	2510	2	T28160	hypothetical protee
13	102	5.9	1132	2	T31107	telomerase reverse
14	101.5	5.9	657	2	AD1525	probable cell surf

15	101.5	5.9	1539	2	S05603	major merizolate su
16	101	5.8	537	2	G64432	hypothetical prote
17	101	5.8	566	2	L19063	hypothetical prote
18	100.5	5.8	1631	1	SAZQKL	major merizolate su
19	100	5.8	978	2	A70387	conserved hypotet
20	99	5.7	679	2	S06000	penicillin-binding
21	98.5	5.7	616	2	G82885	hypothetical prote
22	98.5	5.7	634	2	D71493	probable DNA helic
23	98.5	5.7	834	2	B82940	conserved hypothet
24	98.5	5.7	1587	2	AB2012	hypothetical prote
25	98.5	5.7	4881	2	T18485	hypothetical prote
26	98	5.6	580	2	C81352	lipid export ABC t
27	98	5.6	745	2	H85048	probable transposo
28	98	5.6	954	2	T03750	exonuclease ABC c
29	98	5.6	2819	2	T09080	probable chloroquin
30	97.5	5.6	546	2	T37665	probable t-complex
31	97	5.6	348	2	B67990	hypothetical prote
32	97	5.6	645	2	C84999	ATP-dependent DNA
33	97	5.6	780	2	T21708	hypothetical prote
34	97	5.6	1714	2	E71609	Ser/Thr protein ki
35	96.5	5.6	252	2	H84423	hypothetical prote
36	96.5	5.6	680	2	G95194	penicillin-binding
37	96.5	5.6	685	2	D98061	penicillinbinding p
38	96.5	5.6	1017	2	D90550	veaa-like (mycopla
39	96.5	5.6	2401	2	T28676	thiopyr protein -
40	96	5.5	617	2	H90551	lipoprotein [impor
41	96	5.5	623	3	S67762	hypothetical prote
42	96	5.5	879	2	F81453	DNA-directed DNA p
43	96	5.5	3161	2	T03042	protein HMMP1 -
44	96	5.5	3163	2	AB2233	versinibactin bio
45	96	5.5	3163	2	T17440	probable polyketid

ALIGNMENTS

RESULT 1

hypothetical protein BB28 - Lyme disease spirochete plasmid B/cp26

C;Date: 13-Feb-1998. #sequence_revision 13-Feb-1998 #text_change 09-Jul-2004

R; Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White

; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.

A: Authors: Smith, H.O.; Venter, J.C.

A:Reference number: A70100: MUID:98065943: PMID:9403685
H/ICIE: genomic sequence of a dynamic ribosome binding site, *Escherichia coli* O157:H7

A:Stratug: preliminary: nucleic acid sequence not shown:

A;Molecule type: DNA

A; Cross-references: UNIPROT:050999

C;Genetics:

2

Pred. No.:

Percent Similarity

Query Match:

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

27

55
F
C
C
T

12

1.2	Length:	414
109.50	Matches:	72
36.02%	Conservative:	44
22.36%	Mismatches:	98
6.31%	Indels:	108
2	Gaps:	14

[illegible]

Abstract

[illegible]


```
Db 503 AsnIleAspLeuThrAsnPhcGluIuyswEtwetGlybArgTySer-----Tyr 519
Qy 814 GAGGACAAAGGGGAGGACCTCCACAGCAGCAGGAGAAAGTATGGACACCGAGTGC 873
Db 520 LysValGluIuysLeuThrHisProAsnThrPheAlaSerTyr-----GluAsnSerLys 537
Qy 874 AGTAACTGAGAAATATGCGCATATTA-----TATAGTGTGAGGCTA 918
Db 538 HisAsnLeuGluIuysLeuThrLysAlaLeuLysTyrMetGluAspTyrSerLeuTyr--- 556
Qy 919 AGCCAAAAGACAGGAGTAACTTACTCAAAATGATACTGACAAATACATGACGGTTCAT 978
Db 557 -----AsnIleValValGlu 561
Qy 979 AGTGTGGAACCTCATTTATAAATATA 1005
Db 562 LysGluLeuLysTyrTyrLysAsnLeu 570

RESULT 3
AD1374
Internalin proteins, probable peptidoglycan bound protein (LpxTG motif) homolog lmo2396
C/Species: Listeria monocytogenes
C/Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 09-Jul-2004
C/Accession: AD1374
R/Glaser, P.; Prangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Biochev,
D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.
Science 294, 849-852, 2001
A/Authors: Kretz, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
ok, C.; Schluerke, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A./Title: Comparative genomics of Listeria species.
A/Reference number: AB1077; MUID:21537279; PMID:11679669
A/Accession: AD1374
A/Status: Preliminary
A/Molecule type: DNA
A/Residues: 1-940 <GLA>
A/Cross-references: UNIPROT:O8V4N9; GB:NC_003210; PIDN:CA000474.1; PID:G16411884; GSPDB:
A/Experimental source: strain EGD-e
C/Genetics:
A/Gene: lmo2396

Alignment Scores:
Pred. No.: 1.63 Length: 940
Score: 108.00 Matches: 75
Percent Similarity: 37.03% Conservative: 52
Best Local Similarity: 21.87% Mismatches: 136
Query Match: 6.22% Indels: 80
DB: 2 Gaps: 16

US-10-617-962-2 (1-1008) x AD1374 (1-940)
Qy 112 ACAGCATCATGGAATGAATATGAGCT-----AGAAATATATATGACTTAC 162
Db 355 ThrValThrTyrLysAlaLysPheThrAlaThrSerSerLysGlyValProLeuSerTyr 374
Qy 163 GCTTGGCTGTAAGTGGTATTCATATGATATGATTAACCTCCGATGAC----- 210
Db 375 SerIleAsnValSerGlnProIleAsnValSerGlnGlnThrAspSerThrValSerVal 394
Qy 211 TATTATGAATTAAGAGAGCTGCGAGAAATTTTCAAGATATATGATATGCTATCTTCA 270
Db 395 PheTyrGlnAspGluAsnGlyAsnGluLeuAlaProThrGlnThrLeuSerGlyLysSer 414
Qy 271 TCTGCATATTAAGTGAAGAAATGATGATCAAAATTTTAAAGATATGCAAAATGTTTAT 330
Db 415 -----GlyGluAspTyrGlnThrThrGluLysThrIleAlaAsn----- 427
Qy 331 AAGATGAACAGCTGATTTGAAGTCAATATCTCAAAACATTGGAAGTTCCTGAGCT 390
Db 428 -----TyrGlnLeuLysGluIle 433
Qy 391 GAAATTAACATGATGCTTATTCAGATGACGATTAATATATGACATATATTTT--- 447
```

```
Db 434 GluIuGlnAlaSerGlyGlnPheThrAspThrAspSerThrValThrTyrValTyrGlu 453
Qy 448 -----TTCTCTGTACAGAAATTCACATCGAGGAGAAATCAACAACTA 489
Db 454 LysAlaAspGlyAlaProValThrValLysTyrAlaAspAlaAspGlyAsnLeuAla 473
Qy 490 AATCCCGCAGATTTTAAATTAATGATTTCTTAATTAATTAATTAATTAATTAATTAATTA 549
Db 474 ThrSerAspThrLeuAsnGlyLysIleAspAlaProTyrGlnThrSerAlaLysSerLeu 493
Qy 550 CTGGAGAGAGAGATTTTTCAAAACCTTTTCAATGATTA-----GAGCTTAA 600
Db 494 SerGlyTyrThrValLysThrThrProAsnAlaThrGlyValPheThrAsnSerLys 513
Qy 601 TCATTAAGATTTATTTAGAGAGAAAAAATTTCTTAACCTTCTTCTTGACACCGAG 660
Db 514 GlnThrValThrTyrValTyrGluLysAlaAspGlyAlaProVal-----ThrVal 530
Qy 661 AGATTACCTGATGACAGATAGTATTGCT-----GACCAACA 702
Db 531 LysTyrValaAspGlyAspGlyAsnGluLeuAlaThrSerAspThrLeuAsnGlyLysIle 550
Qy 703 GAACCGCTTAATGAGAGAGTAAAGACTTAAAGAACTTAAAAAT----- 744
Db 551 AspAlaPro-----TyrGlnThrThrAlaLysSerLeuSerGlyTyrThrValLysThr 569
Qy 745 ---AACAAATCAGATGATTTCTTAATTAATGAGAGCGCTGCAAAACAAAGTATGT 801
Db 570 ProAsnAlaThrGlyValPheThrAsnSerLysGlnThrValThrValTyrValTyrGlu 589
Qy 802 TCA-----TTTATTAAGAAGGTACAA----- 822
Db 590 LysAlaAspGlyAlaProValThrValLysTyrValaAspAlaAspGlyAsnGluLeuAla 609
Qy 823 -----AAGGTAACCGCTCA-----CAGACAGCAGCAAAAGTAT 858
Db 610 ThrSerAspThrLeuAsnGlyLysIleAspAlaProTyrGlnThrThrAlaLysSerLeu 629
Qy 859 GGTACAGCAGTGGCAGTACCTGGAATAATTCGCAATATTTATATAGTGTGAGGCTA 918
Db 630 -----SerGlyTyrThrValLysThrThrProAsnAlaThrGlyValPheThr 646
Qy 919 AGCCAAAAGACAGGATGATCTTACTCAAAATGATGACAAATACA---ATGACGGT 975
Db 647 AsnSerLysGlnThrValThrTyrValTyrGluLysAlaAspGlyAlaProValThrVal 666
Qy 976 CATAGTGT 984
Db 667 LysTyrVal 669

RESULT 4
T88317
ORF MSV156 hypothetical protein - Melanoplus sanguinipes entomopoxvirus
C/Species: Melanoplus sanguinipes entomopoxvirus
C/Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 09-Jul-2004
C/Accession: T88317
R/Alonso, C.L.; Tulman, E.R.; Lu, Z.; Oma, E.; Kutish, G.F.; Rock, D.L.
J. Virol. 73, 533-552, 1999
A/Title: The genome of Melanoplus sanguinipes entomopoxvirus.
A/Reference number: Z20484; MUID:99102612; PMID:9847359
A/Accession: T88317
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-1127 <AFO>
A/Cross-references: UNIPROT:O9VVT6; EMBL:AF063866; NID:G4049647; PIDN:AA097677.1; PID:G4
C/Genetics:
A/Note: MSV156

Alignment Scores:
Pred. No.: 1.65 Length: 1127
Score: 108.00 Matches: 72
Percent Similarity: 38.93% Conservative: 44
```



```

Oy      847 -----GGGAAAGATTTGTACGCCCGTGCCTAATCGGAAAAATTCCGAAT   897
           |||
           :||
           :||
Db      1360 ProValPProAlaAlaAlaAlaAlaAlaIAserylsergly-----    1373
           |||
           :||
           :||
Oy      898 AATTATATAGTGTCGAGGCTAGCCAAAAGACAAGGTAACCTTACTCAAAATGATACT   957
           |||
           :||
           :||
Db      1374 -----SerAlaAlaThrThrcylgluAluAlaAlaThrThrValAlaIAserSer   1390
           |||
           :||
           :||
Oy      958 GACAACT   963
           |||
           :||
           :||
Db      1391 Aspasn   1392

RESULT 6
T41912
structural phosphoprotein - human herpesvirus 7 (strain J1)
C:Species: human herpesvirus 7
A:Variety: strain J1
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C:Accession: T41912
R:Nicholas, J.
submitted to the EMBL Data Library, December 1995
A:Description: Determination and analysis of the complete nucleotide sequence of human h
A:Reference number: Z22022
A:Accession: T41912
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-755 <NC>
A:Cross-references: UNIPROT:P52519; EMBL:U43400; PIDN:AAC54672.1
A:Experimental source: strain J1
C:Genetics:
I>Note: U11
```

Alignment Scores:				
Pred. No.:	2.97	Length:	755	
Score:	104.50	Matches:	71	
Percent Similarity:	39.00%	Conservative:	69	
Best Local Similarity:	19.78%	Mismatches:	106	
Query Match:	6.02%	Indels:	113	
DB:	2	Gaps:	18	
US-10-617-962-2 (1-1008) x T41912 (1-755)				
QY	13	TTAACACCTGATGATAGAGTGGATATCCACCCGTTGAAAGCAATAGCAGAGATATA	72	
DB	109	LeuThrSerPheAsnIysValGlyPheGluThrGluPheIysAsnValAlaIysPheLeu	128	
QY	73	GTACGTATACATAAACTTTTAAAGCAACAGATGAGGCTCATACAGCATCTATAGAAATTCAA	132	
DB	129	LeuIysLeuGlySerPheLeuValGrrpGly-----ThrValThrIleAlaIaIaPsr	145	
QY	133	TAT-----	CGAGCTAAG	144
DB	146	TyrValAsnLeuThrThrGluGluAlaGluIleGlyGluAsnLeuGlnIysAlaIys	165	
QY	145	AAATATATATTAGCTTACGCTTTGGCTGTAAGTGTATTCATATAGTATCTTAACTTCCT	204	
DB	166	AsnAsnMetLeuSerPhe-----ThrIleTyrGlnIleVal	177	
QY	205	GATGACATCTTTATAGAAATPAAAGACCTGCTGAGAGAAATTTACAA--GAATATATGCTCT	261	
DB	178	AspProTyrPasnGluAsnGlyTyrTyrValThrAsnIleAsnArgLeuLeuTyrIleuGly	197	
QY	262	AATCTTTCATGCGACTTATTTAGTGGA-----ATGCGATGATCAAAATTCATPAAAGATATG	315	
DB	198	AsnLeuLeuIleThrLeuHisGlySerTrrPheAsnMetGluIysLeuAlaLeuAsnThr	217	
QY	316	GCAAAATGGTTTTTATAGAGTGAACGTGATTTTGAAGGTCAATATCTCAAAAACATTTGG	375	
DB	218	IleAsn-----	219	
QY	376	AATGTTCTCGAGCTTGAATAATTAACCATTTGAGTCTTTTTCAGATACGATTAATTTATA	435	
DB	220	-----GluIysIysAsnAlaIleLeuIysAlaIleGluAsnAspIysAsnPheVal	236	

[illegible]

```

RESULT 7
CS5079
hypothetical protein AT4g08050 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 16-Feb-2001
C:Accession: CB5079
R:Anonymus, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Spring
Nature 402, 769-777, 1999
A>Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.
A:Reference number: AB5001; MUID:20083488; PMID:1061798
A:Accession: CB5079
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-1428 <STO>
A:Cross-references: GB:NC_001268; NID:G7267445; PIDN:CAB81142.1; GSPDB:GN00140
C:Genetics:
A:Gene: AT4g08050
A:Map position: 4

Alignment Scores:
Pred. No.: 3.36 Length: 1428
Score: 104.00 Matches: 78
Percent Similarity: 35.04% Conservative: 59
Best Local Similarity: 19.95% Mismatches: 156
Query Match: 5.99% Indels: 98
DB: Gaps: 16

US-10-617-962-2 (1-1008) x CB5079 (1-1428)

QY 16 ACACCTGATGATGAAAGTGGATATCCACCCGTTGAAAGCAATA-----GCA 63
||| :|||:|||||
750 ThrValGluGlnArgAsnTyrProLeuGluGlnIysGluSerValTyrPheGluGluIlys 769

```

```
QY 64 GAGATATATGTCGTAATCTTAACCTTTAAGCAACAGATGAGGCTCATACAGCATCAT 123
  |||
  |||
  |||
Db 770 GilyilegiuYrSerAlaAlaAsnPheserArg-----GluAsnIaGluYr 785
  |||
  |||
  |||
QY 124 GGAATTAATATCGAGCTTAAGAAAATTAATATGATTAAGCTTTGGCTGCTGATTT 183
  |||
  |||
  |||
Db 786 AsparAspIleArgGluAspYrAlaAspLeuYrSerLeuIleMetSerAsnYr 805
  |||
  |||
  |||
QY 184 CATATGATCTTAACCTTCCTGATGATCTAT-----TATAGAAAT 222
  |||
  |||
  |||
Db 806 SerGlyIleSerSerMetAspProAspYrAsnValAspGluValGluSerIle 825
  |||
  |||
  |||
QY 223 AAAGAGCTGCTGAAGAAATTTATCAAGATATATGCTTAATCTT-----TCATCTGCA 276
  |||
  |||
  |||
Db 826 ArgProArgGluGlnIleValYrGlnSerPheAspGluPheGluYrSerThrAla 845
  |||
  |||
  |||
QY 277 CTATTAAGGTGAATAATGATGATCAATTTCTTAAGAT-----ATGGCAAAATGGTTT 327
  |||
  |||
  |||
Db 846 ArgIleAsnGlnArgArgAlaGluIleAlaArgGlyLysArgAlaMetSerSerArgYr 865
  |||
  |||
  |||
QY 328 TATAAGAAATGAATGATTTTGAAGCTCAATATCCTCAAAACATTTGG---AATGTTCT 384
  |||
  |||
  |||
Db 866 GluLeuIleAspGluAspIleGluThrGluYrGluProGluSerTrpArgGluYr 885
  |||
  |||
  |||
QY 385 GAGCTTGAATAATTAACCATGAGCTTATTCAGATGACGATTAATTTATGACCTATAT 444
  |||
  |||
  |||
Db 886 LysLeuLeuAsnLysPro-----
  |||
  |||
  |||
QY 445 TTTTCTGTCGACGAAATTCACCTGAGGAAATTCACAAATCAATGCCGCAAGATT 504
  |||
  |||
  |||
Db 892 -----AspGluValThrValGluGlu-----TyrIleAspPhe 902
  |||
  |||
  |||
QY 505 TTTTAATTAATGATTTCTTATTATTAC--TTATCTGCTGAATCTCACTGGAAGAG 561
  |||
  |||
  |||
Db 903 PheGluIleAsnAspPheTrpGlyThrArgYrSerGlyThrLeuIleGlnLeu 922
  |||
  |||
  |||
QY 562 ATTTTCAAAAACCTTTTCAATGATTAAGGCTTAATCACTTACAGAT----- 612
  |||
  |||
  |||
Db 923 ArgLeuLeuGluAspValGlnYrLeuPheGluYrCysIleLeuGluThrLeuMetSer 942
  |||
  |||
  |||
QY 613 -----TATATGAGAGAAAACCTTCTTAACCTTTTGCACACCGGAG----- 660
  |||
  |||
  |||
Db 943 TyrProYrProAlaYrTyrLeuGluGluThrIleGluPheLeuSerThrLeuIleValGlu 962
  |||
  |||
  |||
QY 661 -----AGATTACCTGATGCGAGATAGGATTATTTGGCTGGA 696
  |||
  |||
  |||
Db 963 MetYrGluAlaLeuThrAspPheGluLeuAspThrMetGlyLeuGlyPheLeuThrPhe 982
  |||
  |||
  |||
QY 697 CCAACAGAAAGCGCTTAATGAGAGCTGATTTTAAAGAACTTAAATAATCAATCTTAG 756
  |||
  |||
  |||
Db 983 LeuValAspGluGlnArgYrGlnLeuSerIleLysLeu-----GluGluLeu 999
  |||
  |||
  |||
QY 757 AATGATTTTCTTAATATGAGAGGCT----- 783
  |||
  |||
  |||
Db 1000 PheGlyPheProSerGlyLysGlyThrLysProArgPheAspArgGluGluLeuLysAsp 1019
  |||
  |||
  |||
QY 784 -----GCAAAACAA 792
  |||
  |||
  |||
Db 1020 LeuTrpAlaThrIleGlyAsnAsnLeuProLeuAsnSerAlaArgYrLysSerAsnGln 1039
  |||
  |||
  |||
QY 793 AAGTATAGTTCATTATTAAGAAGGTACAAAAGGTAACGCTCCACAGACAGCGGAA 852
  |||
  |||
  |||
Db 1040 IleTrpSerProValIleArgYrPheGlnGlnSerValAlaAsnValPheYrSerLys 1059
  |||
  |||
  |||
QY 853 ---AGTATTGTACCGGAGTGAACCTGGAATAATGCCGATATATTATATAGT 909
  |||
  |||
  |||
Db 1060 GluSerThrGlyThrValSerAsnThrAsnMetGluMetIleAspYrAlaLeuIleGly 1079
  |||
  |||
  |||
QY 910 GTGAGCTTAAGCCAAAGACAGGGAACCTTACTCAAAATGATATCTGACAAATACA--- 966
  |||
  |||
  |||
Db 1080 IleLeuArgArgThrLysGlyLys---AsnValLeuArgGlyAspLeuAsnAsnLysPro 1098
  |||
  |||
  |||
```

```
QY 967 -----ATGACGGTTCATAGCTTGA 987
  |||
  |||
  |||
Db 1099 ProLeuMetProLeuLeuIleIleLeuCysGly 1109
  |||
  |||
  |||
RESULT 8
T09079
probable chloroquine resistance protein CQ2 (strain 768) - malaria parasite (Plasmodium
C;Species: Plasmodium falciparum
C;Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 09-Jul-2004
C;Accession: T09079
R;Su, X.Z.; Kirkman, L.A.; Fujioaka, H.; Wellens, T.E.
Cell 91, 593-603, 1997
A;Title: Complex polymorphisms in an 330 kDa protein are linked to Chloroquine-resistant
A;Reference number: Z16556; MUID:98054002; PMID:9393853
A;Accession: T09079
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-2708 <SUX>
A;Cross-references: UNIPROT:O15791; EMBL:AF030692; NID:G2642513; PIDN:ANC47853.1; PID:92
A;Experimental source: strain 768; from Brazil
C;Genetics:
A;Gene: c92
C;Keywords: toxin resistance

Alignment Scores:
Pred. No.: 3.49 Length: 2708
Score: 104.00 Matches: 75
Percent Similarity: 35.07% Conservative: 53
Best Local Similarity: 20.55% Mismatches: 115
Query Match: 5.99% Indels: 122
DB: Gaps: 19

US-10-617-962-2 (1-1008) x T09079 (1-2708)
QY 79 ATACTAATCTTTAAGCAACAGATGAGGCTCATACAGCATATGCAATTAATGCA 138
  |||
  |||
  |||
Db 1792 LeuIleAsnPheAsnIleThrAsnAsnAsnValSerGlnAsnYrAsn----- 1807
  |||
  |||
  |||
QY 139 GCTAAGAAATATATATTAAGCTTACCGCTTGGCTGAATGATTCATATATGTAATA 198
  |||
  |||
  |||
Db 1808 -----AspLysIleSerAsp 1812
  |||
  |||
  |||
QY 199 CTTCCTGATGACTATTAATTAAGATTAAGAGCTGCTGAGAAATTTATCAAGATATATG 258
  |||
  |||
  |||
Db 1813 LeuAsnYrThrLysAsnLysAsnGluGlu-----ArgSerPheTrp 1826
  |||
  |||
  |||
QY 259 TCTAATCTTATCTGACATTAATGATGAAGATTAAGTGAATTTCTTAAGATATGCA 318
  |||
  |||
  |||
Db 1827 SerSerSerLeuLysLysLeuLeuThrLysPheAspGluGluIle----- 1841
  |||
  |||
  |||
QY 319 AATGATTTTATTAAGATGAATGATTTTGAAGGTCAATATCTCAAAACATTTGGAAT 378
  |||
  |||
  |||
Db 1842 -----PheSerThrAsnAspLeuArgValPheGluGluYrGluMetPheIleSerAsn 1859
  |||
  |||
  |||
QY 379 GTTCCT-----GAGCTGAAATTAACCATTAAGTGTCTTATTCAGATGAGATTA 429
  |||
  |||
  |||
Db 1860 IleLysYrIleLeuLysMetLysAsnLysIleIleSerSer----- 1873
  |||
  |||
  |||
QY 430 TTATTAAGCATATATTTTCTCTGTACAGAAATTCACGAGAGAAATCAACAATCA 489
  |||
  |||
  |||
Db 1874 -----GluValIlePhePheSerProYrPheLeuPro----- 1884
  |||
  |||
  |||
QY 490 AATGCCGAGATATTTTAAATTAATGATTTTATTTATTTACTTATCTGCT--GTAAC 546
  |||
  |||
  |||
Db 1885 -----ThrValLeuYrAsnLeuPheGluPheLeuArgThrLeuGlyValIlePheThr 1902
  |||
  |||
  |||
QY 547 TCATCGGAGAGAGAGATTTTCAAAAACCTTTTCAATGATGATTAAGGCTTAATCATTA 606
  |||
  |||
  |||
Db 1903 LeuLeuGlyArg-----LeuLysAsnAspYrThrAspIleAsnLeuArgAsnAsp 1919
  |||
  |||
  |||
QY 607 GAGAT-----TATATGAGGAAATAAACTTTCTTAACCTTTCTT 648
  |||
  |||
  |||
Db 1920 ArgAsnIleCysAsnValPheGlnIleThrAlaLysLysAsnSerLysSerTrpAsp 1939
  |||
  |||
  |||
```

Qy	649	CGACGACGGCAGAGATTACCTGATGGACGA---ATAGGTTATTGGCGTGGACCAACGAA	705
		::: :::	
Db	1940	AsnGIuPProAsn-----ProGIuValIysAsnMetAsnThrAspGIuAsnAenThrThr	1957
Qy	706	GCGCCTAA-----	714
Db	1558	ThrThrIysIysIysIysAspAspAsnAsnGIuAsnAspArlIeTyrlIeIleu	1977
Qy	715	---TGG-----AGATGAGTTTAAAGAACTTAAATAACAAATCTAGG	756
		::: :::	
Db	1978	IleTrpAsnIleTyAsnValArgThrIeuTyrlThrGIuArgIeAsnAsnAspArgIys	1997
Qy	757	AATGAGTTTCTAATATGGAAGGGCGTGCAAACAAAGTCTACTCTTT-----	807
		:::	
		:::	
Db	1998	-----AlaIysArgIysTyrlGIuThrPheHisIeThrAsn	2011
Qy	808	-----ATAAAGAGGTA	819
Db	2012	MetAspArlIePheAsnAspAsnAsnCysIleAsnIleIleAsnValGIuAspAsn	2031
Qy	820	CAAAAGGGTAACGCTCCACAGACAGCGAAAGATTTGGTACAGCCAGTGCAGTAAC	879
		::: :::	
Db	2032	LysGIuGIuAsnIleIysAspIeulysTyrlIysIysIeulysThrAsnGIuGIuIlys	2051
Qy	880	CTGGAATA-----TTGCCGATATATTATATAGTGTAGCGCTAACCAA	924
Db	2052	ValAspAsnGIuPheIleGIuValThrAspAsnAsnIleIleGIuIleAsnProIyIys	2071
Qy	925	AAAGACAGGGTAACTTACTCAAAATGATCTGAC--AATCAATGACGGTTCTATAGT	981
		:::	
Db	2072	Lys-----ThrserThrGIuAsnGIuGIuPProAsnIleAsnThrIleAsnGIu	2088
Qy	982	GTTGGAATCATTAAT	996
Db	2089	AsnGIuAsnMetIyr	2093

```

RESULT 9
H97194
uncharacterized ABC transporter, ATPase component CAC2392 [imported] - Clostridium acetobutylicum
C/Species: Clostridium acetobutylicum
C/Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 09-Jul-2004
C/Accession: H97194
J/Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A/Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clostridium
A/Reference number: A96900; MUID:21359325; PMID:21359325
A/Accession: H97194
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-579 <KOR>
A/Cross-references: UNIPROT:O97GH5; GB:AE001437; PIDD:AAK80347.1; PID:g15025405; GSPDB:G:
A/Experimental source: Clostridium acetobutylicum ATCC824
C/Genetics:
A/Gene: CAC2392

Alignment Scores:
Pred. No.: 3.81 Length: 579
Score: 103.00 Matches: 70
Percent Similarity: 37.68% Conservative: 63
Best Local Similarity: 19.83% Mismatches: 110
Query Match: 5.94% Indels: 110
DB: 2 Gaps: 16

US-10-617-962-2 (1-1008) x H97194 (1-579)

```

QY 10 CAATTAACCTCGATGATGAGTGGATATCCACCCGTTGAAAGCAATATGACGAGAT 69
 |||||
 ::::|||||
 Db 192 GlnaenthrValGlyIubeuenglyTy-----IleGlucInulleIleGlyAamglu 209
 |||||
 ::::|
 QY 70 ---ATGAGTCGATCTAACTTAAGCAACAAGATGAGGGTCATACAGATCATATGGA 128
 |||||
 ::::|

Db	210	LysIleValLysLeu-----PheGly	216
QY	127	ATTGAAATATCGAGCTAAGAAAATAATATTAGCTTACGCTTTGGCTGTAAAGTATTATTCAT	186
Db	217	TyrGluLysArgAlaGluLys-----LysPheSerGluIleAsnSerArgLeuYr	233
QY	187	AATGATATCTAAACCTCCGTAGTACTATTATTAAGAATTAAGACAGCTCGTAGACAAATTTAT	246
Db	234	AsnCySglGlnLeuAlaGlnPheTyrSerSerLeuThrAsnProSerThrArg-----	251
QY	247	CAAGATATATGTCTTAATCTTTTCATCTGCACTGATTAAGTGAAAAATGTGATCAAAATTTCT	306
Db	252	-----PheValAsnAsnIleThrTyrValLeuValGlyAlaValGlyLeuLeuAla	265
QY	307	-----AAAGATATGCGCAATGCGTTTATTAAGAATGAACCTGATTTGAAAGTCAATAT	360
Db	270	ValLeuSerGlyLeuSerIleGlyThrIleSerSerPheLeuThrTyrSerThrGlnPhe	289
QY	351	CCTGAAACCATTTGATGATGTCCTGAGCTTGAAATAATACATGATGAGCTTATTCAGAT	420
Db	290	SerGlnProIleAsnAsnValThrGlyValAlaThrGlnLeuGlnAlaAlaPheAlaSer	309
QY	421	GACGATTAATTAATTTATGACCTATATTTTCTCTGTAACAGAAATTCACCTGAGAGAAAT	480
Db	310	AlaGluArgValPheSerIle-----LeuAspGluIlePro-----GluLys	323
QY	481	CAACATCAATAATGCGCGAAGATTTTTTTAATTAATTAATGATTTCTTATTACTTATCTGCT	540
Db	324	LysAspAsnGluGlyAlaLysLysPheGluIleCySglGlyAsnIleSerPheAsn	343
QY	541	GTAATCTCACTCGGAAGAGATTTTTCGAAAACCTTTACATGATGATTAGAGGCTAA	600
Db	344	ValSer-----PheSerTyrAsnLysLysGlnProLeu-----	355
QY	601	TCATTATGGAATTAAT--ATTGAGAGAAAAAAACCTTTCTAAACCTTTCTTTCGACACCG	657
Db	355	-----IleGluAsnPheSerValAspIleLysLys-----	364
QY	658	CAGAGATTAACCTGATGCGCAGAAATAGGTATTTTGGCTGACCAACAGAGCGCTAAATGG	717
Db	365	-----GlySerThrIleAlaIleValGlyProThrGlyAlaGlyLysThr	379
QY	718	AGAGTG-----	723
Db	380	ThrMetValAsnLeuLeuMetArgPheTyrAspIleAspGlyGlyIleThrIleAsp	399
QY	724	-----AGTTTAAAGAACTTAAATAATTAACAATCTAGAGATGATTT-----	765
Db	400	GlyLysAspIleAsnLysMetLysArgAsnAspValArgGlyGlnPheGlyMetValLeu	419
QY	765	-----	765
Db	420	GlnAspThrTyrLeuPheGluGlyThrIleLysGluAsnIleAlaTyrGlyLysProAsp	439
QY	766	-----TCATAATATGAGAGGCGTGGAACAAACAAGAT--AGTTCATTTATPAAA	813
Db	440	AlaSerMetGluGlnIleGlnSerAlaAlaLysLysValTyrIleHisAsnPheIleLys	455
QY	814	GAGGTACAAAGGTTAACCGCTCCACAGACAGACGCAAAAGTATTTGGTACAGCAGCTGCG	873
Db	460	ArgLeuSerAspGlyTyrAspThrArgIle-----ThrGlnSerGly	473
QY	874	AGTAACCTCGAAAAAATTCGCCGATAATTTATATAGTGTG	912
Db	474	GlyAsnLeuSerGluGlyGlnLysGlnLeuThrIle	486

```

RESULT 10
S60178
gag polyprotein homolog - fungus (Fusarium oxysporum) retrotransposon skippy
C:Species: Fusarium oxysporum
C:Date: 15-Mar-1996 #sequence_revision 01-Mar-1996 #text_change 09-Jul-2004
C:Accession: S60178
R:Anaya, N.; Roncero, M.I.G.

```



```

OY 415 TCAGAT---GACGAAATTAATTAACACATATTTTCTCTGTACAGGAAATTCACATG 471
    ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 650 LysAsnHisAsnGluTyrThreuproAsnPherPheAsnGluAspIleuLys 669
    ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 472 GAGGAAATTCACAAATCAAT-----GCC 495
    ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 670 AsnAspAspPheIleSerAsnAspAspLysMetAsnLeuGluTyrTyrLeuGln 689
    ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 496 GCAAGATTTTAAATTAATGATTTCTATTTACCTATATCGCTGTAATTCATCGGGA 555
    ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 690 AsnLysIlePheLysAsnLysAsnPhelLeu---SerPheGluAlaIleProLysAspThr 708
    ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 556 AGGAGATTTTTCACAAATTTTACATGATGATGAGGCTAAATCAATTCATGAGATTAAT 615
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 709 GluIleIlePheGluLysLysTyrTyrPheGly-----LeuGluLysAspThr 723
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 616 ATTGAGAGAAAAAATTTTAA-----CCTTTCTTTCGACACCGGAGATTAATCT 669
    ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 724 Val---ArgLysArgIleTyrLysSerIleProIlePheHisIleGluAsnValAlaArg 742
    ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 670 GATGGCAATAGTATTATTTGCT-----GGACCA 699
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 743 HisLysTyrTyrAsnTyrGlnAlaValIleSerTyrAspTyrSerCysPheAlaPro 762
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 700 ACAGAAAGCGCTAAATG---AGAGTGAATTTTAAAGAACT----- 738
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 763 AsnGlu-----TrpLysLysSerPheSerGluTyrProProGluPheGlnAsn 779
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 739 -----AAAAAACAATCTAGAAATGATTTTCT 768
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 780 ThrValHisTyrLeuLeuLeuIleAsnAsnLysLysLysAsnGlnAsnValAlaPheCys 799
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 769 AATATGAAAGGGGCTGCAAAACAAAGTATGATCTATTAATAAGAGTACAAAGGCT 828
    ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 800 TyrThrLysAspProAsnLys-----LysAsnCysAsnLysGly 812
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 829 AACGCTCCACAGACAGCAGCAAGAAATGATGCTACAGCAGGAGGAGTACCTGGAATA 888
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 813 AsnLysGluHisLysLysLysAsnValAsnIleLysAsnAsnGluAsnGluAsn---GluAsn 831
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 889 TTGCGCAATTAATTTA-----TATAGTGAAGGCTAAAGCCCAAAAGAC 930
    ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 832 ValAsnValAsnValAsnValAsnGluGluAspTyrLysAspHisPheAsnGlnHisAsp 851
    ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 931 AGGCTAACCTTTACTCAAAATGATCTGAC 960
    ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 852 LysAsnArgTyrAsnLysAsnLysMetAsp 861
    ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 12
T28160
hypothetical protein - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T28160; S23688
R:Kun, J.F.; Hibbs, A.R.; Saul, A.; McCol, D.J.; Corpe, R.L.; Anders, R.F.
Mol. Biochem. Parasitol. 85, 41-51, 1997
A:Title: A putative Plasmodium falciparum exported serine/threonine protein kinase.
A:Reference number: Z20482; MUID:97262159; PMID:9108547
A:Accession: T28160
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2510 <KUN>
A:Cross-References: UNIPROT:Q94658; EMBL:U40232; NID:G1658332; PID:G1658333; PTDN:AA8540
A:Experimental Source: strain FCQ27/PNG
R:Kun, J.; Heselbach, J.; Schreier, M.; Scherf, A.; Gysin, J.; Mattei, D.; Pereira da
Res. Immunol. 142, 199-210, 1991
A:Title: Cloning and expression of genomic DNA sequences coding for putative erythrocyte
A:Reference number: S23684; MUID:91376328; PMID:1896607
A:Accession: S23684
A:Molecule type: DNA
A:Residues: 241,'S',243-244,673-959,'R',961-977,'S',1493-1494 <KUD>
A:Cross-References: EMBL:X53019
C:Genetics:

```

```

A:Gene: FESP
A:Introns: 645/2
A:Keywords: surface antigen

Alignment Scores:
Pred. No.: 4.51 Length: 2510
Score: 102.50 Matches: 83
Percent Similarity: 40.90% Conservative: 63
Best Local Similarity: 23.25% Mismatches: 106
Query Match: 5.91% Indels: 105
DB: 2 Gaps: 20

US-10-617-962-2 (1-1008) x T28160 (1-2510)
OY 82 CTAACTTAAGCAAAAGAGAGGCTGATACAGATCATATGGAATTAATGAGACT 141
    ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2161 IlserTyrLysArgCysAsnGlu-----IleGlnSerPheGly----- 2173
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 142 AAGAAATTAATTAATGATTAATGCTTTGGCTTAAGTGAATTAATGATTAATGCT 201
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2174 ---LysValGluLeuThrIleLysAsnIleGluSerGlyIlePheLysLeu----- 2189
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 202 CCGATGACTATTAATTAAGATTAAGAGACTGCTGAGAGAAATTTATCAAGAAATATGCT 261
    ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2190 -----ArgAsnLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 2204
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 262 AATCTTATGCTGACATTTAGTGAATTAATGCTGATCA----- 300
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2205 AsnValGlnAsnAspMetMetAsnProAsnLysAspTyrTrpHisSerArgLeuThrAsn 2224
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 301 -----ATTTCTAAGATTAAGCA---AATGCTTTTATAAG----- 333
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2225 LysTyrLeuAlaCysIleLeuLysAspIleCysHisAspAspPheTyrAsnArgMetLys 2244
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 334 -----AATGAAGTGAATTTGAAGGCTCAATATCTCAAAACATTTG 375
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2245 LysIleLysLysSerPheAsnLysTyrGluLeuProLeuAsnTyrSerAspGluTyrTrp 2264
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 376 AATGCTCTGAGCTTGAATAATTAACATTAAGCTGCTTATTAATGAGATTAATTAATTA 435
    ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2265 AspLeu-----LeuThrAsn-----LeuLysAsnTyrValProSerGluArgLeuLeu 2280
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 436 GCACTATAT-----TTTTTCT-----GTACAGGAAAT 465
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2281 AlaCysGluValIleGluLysHisAspPheSerGluHisAsnGluLysIleHisAsnIle 2300
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 466 CCACTGAGGAAATTCACAAATCAATGCGCAAGATTTTAAATTAATTAATGATTTCTTA 525
    ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2301 ---IleLysGluAsnAsnHisAlaArgHisValGluPhePheLysAspGluThrPheCys 2319
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 526 TTATCTTATGCTGCTGATCTTCACTGGGAAAGAGATTTTTCAAAACCTTTTACAAT 585
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2320 AspThrLeu----- 2322
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 586 GGATTAAGAGCTAAATCATTAAGATTAATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 633
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2333 -----LeuLysAsnTyrLysGlnGluLysGluLysAsnTyrIle 2336
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 634 TCTAACTTTCTTTCGACACCGAGAGATTAATCTGATGAGAGAGATTAATTTGCT 693
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2337 SerArgProPhePheGlnAspLeuArgLysLysLeuAspGly-----ValTyrGluGln 2354
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 694 GACCAACAGAGAGCGCTAAATGAGAGCTGAGTTTAAAGAACTTAAATTAACAAATCT 753
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2355 -AspGlnLysLysAsnIleGlnSerTyrTyrLysLysLys---LeuLysLysCysSerIle 2373
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 754 AGGATGATTTTCTTAATATGAGAGGCTCAAAAGTATGATGCTTATTAATAA 813
    ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2373 eProLeuGluIle-----IleArgAsnLysIleIleLeuAsnSerLysIle 2388
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 814 GAGTCAAAAGGCTAACGCTCCACAGACAGAGGAAAGATTTGTTGACAGCAGCTGCC 873
    ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2388 eGlnLysLysAsnValIleIle---MetIleAsnValMetIleIleAspThrThrAsnAsn 2407
    ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```



```

QY 136 CGAGTAAAGAAATATATAGCTTACCGCTTGGCTGAAGTGTATTCATATAGTA--- 192
Db 58 -----IleThrLeuAlaAspValGIunThrIleThrSerLeuAspProValGly 73
QY 193 ---TCTAAATCTCTGATGACTAT-----TATAAGATTAAGAAGACTGCTGAGCAATT 243
Db 74 AlaserSerIleProAspAsnIleThrAspTyrIysAsn-----Leu 87
QY 244 TATCAAGAAATATATCTATCTTCACTGACACTATAGGAAATGATGATCAATT 303
Db 88 ThrArgLeuTyrIleThrGlnGlyThrLeuThrGluValProGluSerIleGlyLeu 107
QY 304 TCTAAAGATATGCAAAATGCTTTTATATAGAAATGCAATTTTGAAGTCAATATCT 363
Db 108 LysIleSleuThrPheLeuSerPheTyrAsnAsnIleuThr-----GluPhePro 124
QY 364 CAATAATTTGGAAATGTTCTGAGCTTGAATAAATCAATGATGCTTATTCAGATGAC 423
Db 125 ThrValValTyrAspLeuProAlaLeuAsnSer----- 135
QY 424 GATAAATTTATAGCACTATATTTTCTCTGACAGAAATTCACATGAG---GAAAT 480
Db 136 -----LeuLeuLeuGlnArgAsnAsnIleSerGluIleProGluGluIleThrAsn 152
QY 481 CAACATCAAAATGCCCA-----AGATTTTAAATTAATTAATGAT 519
Db 153 MetSerSerHisLeuSerSerLeuAspValArgAsnAsnAsnIleSerIleProAsp 172
QY 520 TTTCTATTATTTACCTTATCTGCTGTAATTCTCACTGGAAGAGAGATT----- 564
Db 173 LysIlePheThrThrIleTyrAlaSerArgSerGlyGlnLeuIleIleAspThrGluGly 192
QY 565 -----TTTCAAAAAGCTTTTACAAATGATGATTAAG 594
Db 193 AsnGlnIleThrSerAspValProValAspTyrLeuAspAsnTyrAsnAsnGly----- 210
QY 595 GCTAAATCATTAAGAAATTAATATGAGAAATAAATCTTAAACCTTTCTTGACCA 654
Db 211 GlyAsnMetLeuGlnAsnIleThrAsnTyrArgGlnIys----- 222
QY 655 CCGCAGAGATTACCTGATGGCAGAAATAGGTTATTTGGCTGACCAACAGAACGCGCTTAA 714
Db 223 -----GlnAspGlnLeuValTyrIysGlyAspProIleValValPro--- 236
QY 715 TGGAGAGTGAAGTTTAAAGAACTTAAATAAATCAATCTAGAAATGATTTTCTAAT 771
Db 237 TyrIysThrAspPheIleGlnLeuThrProAspLysSerIleuGlyLeuAlaSer 255

RESULT 15
S05603
major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum) (sp
N:Alternate names: gp195 surface antigen
C:Species: Plasmodium falciparum
C:Date: 12-Feb-1993 #sequence_revision 12-Feb-1993 #text_change 09-Jul-2004
C:Accession: S05603; S04850
R:Myler, P.J.
submitted to the EMBL Data Library, April 1989
A:Reference number: S05603
A:Accession: S05603
A:Molecule type: mRNA
A:Residues: 1-1639 <MYL>
A:Cross-references: UNIPROT:P04933; EMBL:X15063; NID:g9896; PIDN:CAA33163.1; PID:g9897
Nucleic Acids Res. 17, 5401, 1989
A:Title: Nucleotide and deduced amino acid sequence of the gp195 (MSA-1) gene from Plasm
A:Reference number: S04850; WUID:89345116; PMID:2668887
A:Accession: S04850
A:Molecule type: mRNA
A:Residues: 1504-1639 <MYL2>
A:Cross-references: EMBL:X15063
C:Superfamily: major merozoite surface antigen
C:Keywords: glycoprotein; merozoite; surface antigen
F:1-19/Domain: signal sequence #status predicted <Sig>

```

```

F:20-1639/Product: major merozoite surface antigen #status predicted <MYL>
Alignment Scores:
Pred. No.: 5.25 Length: 1639
Score: 101.50 Matches: 82
Percent Similarity: 34.70% Conservative: 53
Best Local Similarity: 21.08% Mismatches: 137
Query Match: 5.85% Indels: 117
DB: 2 Gaps: 19
US-10-617-962-2 (1-1008) x S05603 (1-1639)
QY 121 TATGAATTAATGATTCGAGCT-----AGAAATATATATAGCTTAC 162
Db 205 PheAsnLeuIysIleArgAlaAsnGluLeuAspValLeuIysIleValPheGlyTyr 224
QY 163 GCTTTGCTGTAAGTGTATT---CATATGTATCTAAATCTCTGATGACTATATATAG 219
Db 225 ArgIysProLeuAspAsnIleIleAspAsnValGlyIysMetGluAspTyrIleIysIys 244
QY 220 AATAAGGAGCTGCTGAG----- 237
Db 245 AsnIysIleThrIleGluAsnIleAsnGluLeuIleGluIleSerIleThrIleAsp 264
QY 238 -----AGATTTATCAAA---GAATATATG 258
Db 265 LysAsnIysAsnAlaThrIleGluGluGluIleIysIleIysIleValIleAsnGlu 284
QY 259 TCTAATCTTCACTGCACTATTTAGCTGAAAATGCTGATCAATTTCT---AAAGAT 312
Db 285 LeuSerIleTyrAsnIysGlnLeuGluGluIleAsnLeuIleSerValLeuGluIys 304
QY 313 ATGCAAAATGCTTTTAAAGAAATGAA-----CTGGAT----- 345
Db 305 ArgIleAspThrLeuIleIysAsnGluAsnIleIysGluLeuLeuAspIleAsnGlu 324
QY 346 TTTGAGGCTCAATATCTCAAAACATTTGGAATGTTCTCTGAGCTT-----GAAAT 396
Db 325 IleIysAsnProProProAlaAsnSerGlyIleAsnThrProAsnThrLeuLeuAspIysAsn 344
QY 397 AAACCATTAAGTCTTATTCAGATGACATGAATTAATTAATGACATATATTT---TTCTCT 453
Db 345 LysIysIleGluGluIleGluIleGluIleIleIysGluIleAlaIleIleThrIleIysPheAsn 364
QY 454 GTACAGCAAAAT-----CCACTG----- 471
Db 365 IleAspSerLeuPheThrAspProLeuGluLeuGluIleIleIleIleIleIleIleIleIle 384
QY 472 -----GAGAAATATCAACATCA 489
Db 385 AsnIleAspIleSerAlaIleValGluThrIysGluSerThrGluProAsnGluTyrPro 404
QY 490 AATGCCGCAAGATT---TTTAAATTAATGATTTCTTATTTACCTTATCTGCTGTAAT 546
Db 405 AsnGlyValThrTyrProLeuSerTyrAsnAspIleAsnAlaLeuAsnGluLeuAsn 424
QY 547 TCACCTGGAGAGAGATTTT-----TCAAAAACCTTTAA 582
Db 425 SerPheGlyAspLeuIleAsnProPheAspTyrThrIysGluProSerIleIysAsnIleTyr 444
QY 583 AATGATTAAGAGGCTTAATCACTTAGAGATTAATATGAGAGAAA---AACTTTCTTAA 639
Db 445 ThrAspAsnGluIleGluIleGluIleAsnGluIleIysGluIleIleIleIleIleIleIleIle 464
QY 640 CTTTCTTTGACACCGCAGAGATTACCTGATGCGAGAAATAGGTTATTTGGCTGACCA 699
Db 465 LysIysIleGluSerAspIleIysSerTyrGluAspArgSerIysSerLeuAsnAspIle 484
QY 700 ACAGAGAGGCTTAATGAGAGTGAATTTTAAAGCTTAAATAAATAAATACTAGAT 759
Db 485 ThrIys-----GluTyrGluIleIleuLeuAsnGluIleIleTyrAspSerIlePheAsn 502
QY 760 GGA-----TTTCTAATATGGAAGGCGCTGCAAAACAAAGTATATGTTCAATTATTA 813

```

```

Db      503 AsnIleAspLeuThrAsnPheGluLysMetMetGlyLysArgTyrSerTyrLysValGlu 522
QY      814 GAGGTACAAAAGGTAAACGCTCCACAGACAGCAGCGAAAGTATTGTTACAGCCAGTGGC 873
Db      523 LysLeuThrHisHisAsnThrPheAlaSerTyrGluAsn-----SerLys 537
QY      874 AGTAACTGTGAAAAATTGCCGAATATTAA-----TATAGGTGAAGCTA 918
Db      538 HisAsnLeuGluLysLeuThrLysAlaLeuLysTyrMetGluAspTyrSerLeuArg--- 556
QY      919 AGCCAAAAAGACAGGTAACCTTTACTCAAAATGATACACATATCAATGACGGTTCAT 978
Db      557 -----AsnIleValValGlu 561
QY      979 AGTGTGGAACCTCATTTATATAATATA 1005
Db      562 LysGluLeuLysTyrTyrLysAsnLeu 570

```

Search completed: November 21, 2004, 08:49:11
 Job time : 54 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 21, 2004, 06:30:11 ; Search time 3133 Seconds
(without alignments)
11723.986 Million cell updates/sec

Title: US-10-617-962-2

Perfect score: 1 atggttatacaataaacacc.....ctcatataaaatatatga 1008

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 32822875 seqs, 18219865908 residues

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	62.8	6.2	1101	CNS0039G	AL063921 Drosophila
C 2	61.2	6.1	1225	CNS0161D	AL061721 Drosophila
C 3	60.6	6.0	1101	CNS008EV	AL069706 Drosophila
C 4	59.4	5.9	996	CNS000FV	AL071063 Drosophila
C 5	57	5.7	1101	CNS0021J	AL061936 Drosophila
C 6	56.4	5.6	928	CNS000DK	AL071865 Drosophila
C 7	55.8	5.5	1101	CNS000EVL	AL069706 Drosophila
C 8	55.6	5.5	844	BX139987	BX139987 Drosophila
C 9	55	5.5	1101	CNS0039G	AL063921 Drosophila
C 10	54.6	5.4	1084	CNS0071NH	AL063921 Drosophila
C 11	54.2	5.4	561	CAS56457	CAS56457 PESTOACO
C 12	53.8	5.3	1101	CNS016HF	AL06749 Drosophila
C 13	53.6	5.3	924	BX398867	BX398867 Drosophila
C 14	53.4	5.3	1092	CNS020K7	AL175866 Tetraodon
C 15	52.4	5.2	1327	CL644705	CL644705 CH213-79D
C 16	51.8	5.1	1074	CNS015ZR	AL106113 Drosophila
C 17	51.6	5.1	987	CNS014PQ	AL104456 Drosophila
C 18	51.6	5.1	1011	CF469747	CF469747 P4-CS Pia
C 19	51.4	5.1	600	CG91559	CG91559 EST011 E
C 20	51	5.1	1350	CG744271	CG744271 P036-4-C0
C 21	50.8	5.0	853	CNS023KH	AL179594 Tetraodon
C 22	50.6	5.0	489	CF181864	CF181864 ISOIGF I
C 23	50.6	5.0	519	CF181855	CF181855 ISOIF1 I
C 24	50.6	5.0	976	CL516956	CL516956 SATL_98_C

25	50.6	5.0	1101	9	CNS0006J	AL062049 Drosophila
26	50.4	5.0	567	4	BM165529	BM165529 EST568052
27	50.4	5.0	807	9	CL489403	CL489403 SATL_523
28	50.4	5.0	905	8	AZ550256	AZ550256 ENTREV58TR
29	50.4	5.0	1003	6	CD390192	CD390192 AGENCOURT
30	50.2	5.0	1146	9	CNS021G2	AL176643 Tetraodon
31	50	5.0	446	2	BF330152	BF330152 RC0-BN031
32	50	5.0	548	2	BF330145	BF330145 RC0-BN031
33	50	5.0	770	8	BH549511	BH549511 BOGFS24TR
34	50	5.0	832	8	BH391984	BH391984 AG-ND-138
35	50	5.0	928	9	CNS00DKY	AL071865 Drosophila
36	50	5.0	1133	9	CG752293	CG752293 P046-4-H1
37	49.8	4.9	717	8	BH957562	BH957562 odg07E10
38	49.8	4.9	746	8	BH591510	BH591510 BOGDH24TR
39	49.8	4.9	849	8	AZ546009	AZ546009 ENTREV53TR
40	49.6	4.9	461	4	B1424095	B1424095 B1424095
41	49.6	4.9	924	4	CNS02Q6S	AL208909 Tetraodon
42	49.6	4.9	1047	9	CNS01429	AL103611 Drosophila
43	49.4	4.9	1101	9	CNS003BB	AL064089 Drosophila
44	49.2	4.9	521	1	AU264780	AU264780 AU264780
45	49.2	4.9	550	9	CC952969	CC952969 BOIHJ75TR

ALIGNMENTS

RESULT 1
CNS0039G/c 1101 bp DNA linear GSS 03-JUN-1999
LOCUS
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR08K10 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.

ACCESSION AL063921 GI:4941778
VERSION
KEYWORDS
SOURCE
ORGANISM Drosophila melanogaster (fruit fly)
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE
AUTHORS
TITLE
JOURNAL
Genoscope.
Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)

COMMENT
Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org/TheBDGP/Drosophila_melanogaster_BAC_library was prepared by Kazutoyo Osoegawa and Aaron Mammeter in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES
source
location/Qualifiers
1..1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR08K10"
/clone_11b="RPCI-98"
/note="end : TET3"

ORIGIN

Query Match 6.2%; Score 62.8; DB 9; Length 1101;
Best Local Similarity 20.5%; Pred. No. 0.00098;

Dd		862	AAATYMAAYTTAAATTAMAAAATAATWMTTTTTTTTTTTTTTTTTTTTTTTCSSSCSSCSSSSS	921
Oy		470	TGAGAGAAATCATCAACATCAAATGGCCGACAAGATTTTTTAATAATTAATGAATTTCTATTT ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::	528
Dd		922	CSSCVAANAATTAATTAATTAATTAATTTTWTMTWTMTTTKAATAATWMAAAAAATTTTTTT ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::	980
RESULT 5				
LOCUS				
DEFINITION		CNS0021J	1101 bp	DNA linear GSS 03-JUN-1999
VERSION		Drosophila melanogaster genome survey sequence TET3 end of BAC #		
KEYWORDS		BACR05H1 of RPci-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.		
SOURCE		AL061936		
ORGANISM		AL061936.1 GI:4940214		
		GSS.		
REFERENCE		Drosophila melanogaster (fruit fly)		
AUTHORS		Drosophila melanogaster		
TITLE		Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;		
JOURNAL		Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;		
		Ephydroidea; Drosophilidae; Drosophila.		
		1 (bases 1 to 1101)		
		Genoscope.		
		Direct Submission		
		Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :		
		BP 191 91006 EVRY cedex - FRANCE (E-mail : segre@genoscope.cns.fr		
		- Web : www.genoscope.cns.fr)		
		Determination of this BAC-end sequence was carried out as part of a		
		collaboration with the Berkeley Drosophila Genome Project (BDGP).		
		The BDGP is constructing a physical map of the Drosophila		
		melanogaster genome using these BACs. For further information		
		please see http://www.fruitfly.org The BDGP Drosophila		
		melanogaster BAC library was prepared by Kazutoyo Osoegawa and		
		Aaron Mammoler in Pierer de Jong's laboratory in the Department of		
		Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,		
		NY. The library is named RPci-98 and was constructed by partial		
		EcORI digestion of Drosophila DNA provided by the BDGP from the		
		isogenic strain Y2; cn bw sp, the same strain used for the BDGP's		
		P1 and EST libraries. A more detailed description of the library		
		and how to order individual BAC clones, the entire library, or		
		filters for hybridization from the BACPAC Resource Center can be		
		found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.		
FEATURES				
source		Location/Qualifiers		
		1..1101		
		/organism="Drosophila melanogaster"		
		/mol_type="genomic DNA"		
		/db_xref="taxon:7227"		
		/clone="BACR05H1"		
		/clone_1lb="RPci-98"		
		/note="end : TET3"		
ORIGIN				
Query Match		5.7%; Score 57; DB 9; Length 1101;		
Best Local Similarity		37.2%; Pred. No. 0.021;		
Matches		289; Conservative 59; Mismatches 425; Indels 3; Gaps 1		
Oy		50	AAAAGCAATAGCAGAGATATAGCTAATCTTAAGCAACAGATGAGGTC	109
Dd		312	AAAAAACAAA	371
Oy		110	ATACGCATCATGTGAATTCGACTAAGAAATAATATTAGCTTAGCGTTGG	169
Dd		372	AAA	431
Oy		170	CTGTAGTGATTCATATATCTTAACTTCGATGACTATATAGAATAAGACA	229
Dd		432	AAA	491
Oy		230	CTGCTAGAGATTTATCAAGATATATCTTAATCTTCATCTGCATATTAGGAAA	289
Dd		492	AAAAAAAAAAAAAAAAAAAAAAAAATTAATTTTATTTTATTAATTTTTTTTTT	551
Oy		290	ATGGATCAATTCTTAAGATAGGCCAATGGTTTTTATTAAGATGAATGATTTTG	349

[illegible]


```

/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR27A24"
/clone_id="RPC1-98"
/note="end : 17"

```

Query Match	5.6%;	Score 56.4;	DB 9;	Length 928;
Best Local Similarity	29.3%;	Pred. No. 0.028;		
Matches 113;	Conservative 91;	Mismatches 182;	Indels 0;	Gaps 0;

[illegible]

LOCUS	CNS600EWL	1101 bp	DNA	linear	GSS 04-JUN-1998
DEFINITION	Drosophila melanogaster genome survey sequence T7 end of BAC: BACR29923 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.				
ACCESSION	AL069706				
VERSION	AL069706.1	GI:4949849			
KEYWORDS	GSS.				
SOURCE	Drosophila melanogaster (fruit fly)				
ORGANISM	Drosophila melanogaster				
	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephyridae; Drosophilidae; Drosophila.				
REFERENCE	1 (bases 1 to 1101) Genoscope. Direct Submission Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr - web : www.genoscope.cns.fr)				
AUTHORS	determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila genome project (BDGP).				
TITLE	The BDGP is constructing a physical map of the Drosophila melanogaster genome using three BACs. For further information please see http://www.fruitfly.org The BDGP drosophila				
JOURNAL	melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammosser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the Ito strain Y2, cn bw sp, the same strain used for the BDGP's p1 and EST libraries. A more detailed description of the library				
COMMENT					

and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

Location/Qualifiers

1. .1101

Location="Drosophila melanogaster"

FEATURES

source

```

source
1. .1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR29B23"
/clone_jib="RPCT-98"
/note="end : T7"

```

Query Match 5.5%; Score 55.8; DB 9; Length 1101;
Best Local Similarity 31.4%; Pred. No. 0.039;
Matches 201; Conservative 109; Mismatches 328; Indels 3; Gaps 1

QY	184	CATATGATGATTAACCTTCGATGACGATTAATGAAGATTAAGACGCTGAGGAATT	243
Db	446	MMGEMMATYTCCHATTMMGMMGMAATWMMWAAWMAAAATTATATWATWMAAAWMAWMM	505
QY	244	TATCAGAAATATGTCCTAATCTTTCATCTGCACTATTAGTGAAATGCGATCAATT	303
Db	506	WATTTTTTMMWWTATTTTWTWMMWTWMTAAAAAATAATATTTAAAAWMAATWMT	565
QY	304	TCTAAGATATGCGAAATGCTTTTATTAAGAATGAACCGAATTTGAGAGCTAATCCT	363
Db	566	TAAWAAATTTAAWAAWMTATATTAATWTAATAATWATTAATTAATAAAATATTTTTTWA	625
QY	364	CAAAACATTTTGAATGTTCCGAGCTTGAAATTAACCATGAGTGTATTCAGATGAC	423
Db	626	TAAATTTTTTATATTAATTAATTWATTAATAATWMAATTTMMWTAAATTAATTAATTTWA	685
QY	424	GATTAATTTATGACATATATTTTTTCTCTGTGACGAAATTCCTGAGGAAATCA	483
Db	686	AATWMAAAAAAAAAAAAAAAAAWMAWMAATATATATATTTAAWMAATAAAWMAWMA	745
QY	484	CAATCAAAATGCGCAAGATTTTAAATTAATGATTTCTATTAACCTTATCTGCTGTA	543
Db	746	TWAAWATATATATWATATATATTTTAAWMAATWMAWMTATATMAATTAATAAAWMA	805
QY	544	ACTTCACTGGAGAGAGATTTTTTCAAAAAACITTTACAATGATTAAGAGCTAAATCA	603
Db	806	---TAAATAMATWMAWMAAAWMAWMAWMTATATWMAATTAWMAAAAAATWTAATA	862
QY	604	TTAGAGATTTATTTGAGAGAAAAAAACITTTCTAAACCTTCTTGACACCCGAGAGA	663
Db	863	TWATWMAWMAAAAAAAAAATWMAWMTTWTTTTTTWWMAWMTATATAAAWMAWMAWMAA	922
QY	664	TTACCTGAGCAGAAATAGTTATTGTGCTGGAACAACAGAGAGCCCTAAATGAGAGATG	723
Db	923	AAAAAAAAATTAAMAWMTATATTTWTATTTAAATMTATATWATTTWATTTWMAATWTATWTT	982
QY	724	AGTTTTTAAGACTTAAAAATTAACAAATCTAGAAATGATTTTCTAATATGGAAGGGCT	783
Db	983	WTATWTATATATWMTATWMAWMTATATTTATTAAMWTATATTTTAAAAWMTAATATATWAT	1044
QY	784	GCAAAACAAAAGTATGTTCAATTTTATAAAGAGTACAAA	824
Db	1043	AMWMTAMATATAMWMAATTTWTTTATATATATWTAATAATTAATA	1093

RESULT 8	844 bp	DNA	linear	GSS 13-MAR-2001
EX133987				
LOCUS				
DEFINITION	Danio rerio genomic clone DKEX-99E7, genomic survey sequence.			
ACCESSION	EX133987			
VERSION	EX133987.1	GI:27971314		
KEYWORDS	GSS.			
SOURCE	Danio rerio (zebrafish)			
ORGANISM	Danio rerio			
	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi;			

REFERENCE
AUTHORS
TITLE
JOURNAL

Cypriniformes; Cyprinidae; Danio.
1 (bases 1 to 844)
Humphray, S.J., Huckie, E. and Durham, J.L.
Direct Submission
Submitted (13-MAR-2003) The Sanger Institute, Wellcome Trust Genome
Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humphray@sanger.ac.uk unpublished
This sequence was generated from the T7 end of BAC 99E7. 99E7 is
part of the Dantolkey BAC library created by R. Plasterk and N.V.
Keygene. Further details:
http://www.sanger.ac.uk/Projects/D_danio/.

COMMENT

FEATURES
source

location/Qualifiers
1..844
/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/clone="MDKEY-99E7"
/rissue_type="Testis"
/note="vector pindigobac-536"

ORIGIN

Query Match 5.5%; Score 55.6; DB 9; Length 844;
Best Local Similarity 51.2%; Pred. No. 0.042; Mismatches 124; Indels 0; Gaps 0;
Matches 130; Conservative 0;

QY 175 AGTGGATTCATTAATGATGTAACCTTCCTGATGATCTATTAAGAAATGAAGAGAGAGCTCT 234
DB 359 AATGATATATTAATGATATTAATGATATGATATGATATTAATTAATGATATAT 418
QY 235 GAGGAAATTTTCAAGATATATGCTCTATCTTCATCTGACATTTAGGAAATAGT 294
DB 419 GATATATTAATTAATTAATGATATTAATTAATTAATTAATTAATTAATTAATTAAT 478
QY 295 GATCAATATTTCTTAAGATGCAATAGCTTTTAAAGATGAAGCTGATTTGAAGT 354
DB 479 AATGATATATTAATTAATGATATGATATGATATTAATTAATTAATTAATTAAT 538
QY 355 CAATATCTCTCAAAATTTGAATGCTTCGAGCTTGAATTAATCAATGAGCTTAT 414
DB 539 AATATGATATTAATTAATGATATTAATTAATTAATTAATTAATTAATTAATGATAT 598
QY 415 TCAGATGCGATTA 428
DB 599 AATAATGATGATTA 612

RESULT 9

CNS0039G 1101 bp DNA linear GSS 03-JUN-1999
LOCUS
DEFINITION
Drosophila melanogaster genome survey sequence TET3 end of BAC #
BACR08K10 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION
AL063921.1 GI:4941778
VERSION
GSS.
KEYWORDS
Drosophila melanogaster (fruit fly)
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 1101)
REFERENCE
AUTHORS
TITLE
JOURNAL

Genoscope.
Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 Evry cedex - FRANCE (E-mail : sequefgenoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org The BDGP Drosophila
melanogaster BAC library was prepared by Kazuo Osoegawa and
Aron Mammoser in Pieter de Jong's laboratory in the Department of

Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain Y2; cn bw sp, the same strain used for the BDGP's
P1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES
source

location/Qualifiers
1..1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR08K10"
/clone_lib="RPCI-98"
/note="Tend : TET3"

ORIGIN

Query Match 5.5%; Score 55; DB 9; Length 1101;
Best Local Similarity 18.8%; Pred. No. 0.06; Mismatches 303; Indels 3; Gaps 1;
Matches 132; Conservative 265;

QY 95 AAAGAGAGAGGCTCATCAGCATCATATGGAATGATATGAGAGAGAGAGATTAAT 154
DB 401 WAMMMWTTTTTTTTTAAAMAAAATAATTTWMAAMAAAATTTWMAAMAAAAT 460
QY 155 TAGCTTACGCTTGGCTGTAAGTGATTCATTAATGATTAATCTTCTGATGAT 214
DB 461 AMTTTTTAAAMAAAATAATTTTATTTTATTTTATTTATTTATTTTATTTTAA 520
QY 215 ATTAAGATTAAGAGAGCTGCTGAGAGATTTTCAAGATATATGCTATATCTTGCATG 274
DB 521 AAAAAAAMAAAAMAAAATAATTTTATTTTATTTTATTTTATTTTATTTTATTT 580
QY 275 CACTATTAAGTGAATGATGATCAATTTCTTAAGATGATGCAATAGCTTTTAA 334
DB 581 HTTWTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTAT 640
QY 335 ATGAAGCTGATTTTGAAGCTCATATCTCTCAAAATTTGGAATGCTTGAAGT 394
DB 641 YUUCMTYUHNMMNNNNNAAAMWTTTHTTTHATYUHNATYUHNATYUHNATYUHN 697
QY 395 ATAAACATGAGCTTATGATGATGATGATGATGATGATGATGATGATGATGAT 454
DB 698 CHNCYUHNATYUHNATYUHNATYUHNATYUHNATYUHNATYUHNATYUHNATY 757
QY 455 TACAGAAATTCACATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGATTT 514
DB 758 YAMAMMMNNNNNAAAAMAAATTTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 817
QY 515 TTGATTTCTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 574
DB 818 YUHNATYUHNATYUHNATYUHNATYUHNATYUHNATYUHNATYUHNATYUHNAT 877
QY 575 ACTTTTCAATGATGATGAGGCTTAATCATTAAGATTAATGATGAGAGAGAGAGAG 634
DB 878 AATTTTNN 937
QY 635 TTAACCTTTCTTCTGACACAGAGAGATTAATGATGAGAGAGAGAGAGAGAGAG 694
DB 938 HNNMMATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 997
QY 695 GACCAAG 754
DB 998 HCTTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 1057
QY 755 GGAATGATTTTCTAATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 797
DB 1058 YUHNATYUHNATYUHNATYUHNATYUHNATYUHNATYUHNATYUHNATYUHNAT 1100

RESULT 10

CNS071NH/c

LOCUS	1084 bp	DNA	linear	SSS 07-JUL-2001
DEFINITION	clone BA0A003E09 of library BA0AB from strain CLIB 210 of Kluyveromyces lactis, genomic survey sequence.			
ACCESSION	AF425139.1 GI:12208333			
VERSION	GSS.			
KEYWORDS	Kluyveromyces lactis			
SOURCE	Kluyveromyces lactis			
ORGANISM	Kluyveromyces lactis			
REFERENCE	Bukaryota Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Kluyveromyces.			
AUTHORS	Souciet, J.-L., Aigle, M., Artiguenave, F., Blandin, G., Boivin-Fukuhara, M., Bon, E., Brotier, P., Casaregola, S., de-Montigny, J., Dujon, B., Durren, P., Leplingle, A., Lorente, B., Maupertuy, A., Neuvéglise, C., Ozier-Kalogiropoulos, O., Potier, S., Saurin, W., Tekala, F., Toffano-Nioche, C., Wesolowski-Louvel, M., Winkler, P., and Weissenbach, J.			
TITLE	Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies			
JOURNAL	FEBS Lett. 487 (1), 3-12 (2000)			
MEDLINE	20584711			
PUBMED	11152876			
REFERENCE	2 (bases 1 to 1084)			
AUTHORS	Bolotin-Fukuhara, M., Toffano-Nioche, C., Artiguenave, F., Duchateau-Nguyen, G., Lemaire, M., Marmeisse, R., Montrocher, R., Robert, C., Termier, M., Winkler, P. and Wesolowski-Louvel, M.			
TITLE	Genomic exploration of the hemiascomycetous yeasts: 11. Kluyveromyces lactis			
JOURNAL	FEBS Lett. 487 (1), 66-70 (2000)			
MEDLINE	20584721			
PUBMED	11152886			
REFERENCE	3 (bases 1 to 1084)			
AUTHORS	Genoscope.			
TITLE	Direct Submission			
JOURNAL	Submitted (08-SEP-2000) Genoscope - Centre National de Séquençage, 2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail : beqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)			
COMMENT	This GSS is part of a random genomic sequencing program of thirteen yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii, Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila, Candida tropicalis and Hansenula lipolytica. Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert.			
FEATURES	Location/Qualifiers			
SOURCE	1..1084			
	/organism="Kluyveromyces lactis"			
	/mol_type="genomic DNA"			
	/strain="CLIB 210"			
	/variety="lactis"			
	/db_xref="taxon:28985"			
	/clone="BA0AB003E09"			
	/clone_1id="BA0AB"			
ORIGIN				
Query Match	5.4%	Score 54.6;	DB 9;	Length 1084;
Best Local Similarity	40.7%	Pred. 0.074;		
Matches 183;	Conservative 29;	Mismatches 236;	Indels 2;	Gaps 11.
Db	165	TTTGCGCTTAAGTGGTATTCATTAAGTATGCTAAACCTTCCTGATGACTATTAAGAAATPA	224	
Oy	1035	TTTATTATTTAAATTAAATTAAAAAATTTTWTTTTWTACCTTTTNTNNAAAAATTT	976	
Db	225	AGAGACTGCTGAGAGAAATTTATCAAGAAATATATGTCTAATCTTTCATCTGACATTAAGG	284	
Oy	975	AAATANNATTTTAAAAATWAAAAAAATTTTAAAAATTTTWTTTTATATTTTAT	916	
Db	285	TGAAATAGTGGA--TCAAATTTCTAAAGATATNGCAAAATGCTTTTATTAAGATGAACG	342	
Oy	915	NATATATATATTAATTAATTCATTTATTTATATATTAATTTTATTTTATTAATTTAAATTA	856	

QY	343	GATTTCAGAGTCACATTCCTCCGAAAAACATTTGGAAATCTTCGAGCTTGAAAAATPAAACA	402
Db	855	AATTATTTAAAAATATMTATTTTWTTRAAATATATATATATTTTATATTTTNTTTAAAAAAAAT	796
QY	403	TTGAGTCCTTATTCAGATGACGATTAATTTATTTGACCTATATTTTTCTCTGACAGAA	462
Db	795	TAAATTTTWWMTATATATATATATPAAAAATNTWANAATTTATTTTWTWAMATA	736
QY	463	ATTCCACTGGAGAGAAATCAACATTCAAATCCGCAAGATTTTAAATPATGATTC	522
Db	735	AMTTTTTTTAAATTAATAATATATTTTAAATTTTAAATTTTATTTATTTATTTTATW	676
QY	523	TTATTTACCTATCTGCTGTAACCTCACTCGGAAGAGAGATTTTTCAAAAAACTTTTAC	582
Db	675	AATTTTATTTATTTTATTTTATTTTATATATATATAAAAAAAATAAATTTATTTATTTATW	616
QY	583	AATGATTAGAGCTTAATCATTTAGAGAT	612
Db	615	ATATATAGKKGKRGKRWMAAAAAATTTTATW	586
RESULT 11			
Ca856457		561 bp	mRNA
LOCUS			linear
DEFINITION	PESTscoc8e04.y1 Plasmodium falciparum 3D7 gametocyte cDNA library		EST 17-DEC-2002
ACCESSION	Ca856457		
VERSION	Ca856457.1	GI:27158994	
KEYWORDS	EST.		
SOURCE	Plasmodium falciparum 3D7		
ORGANISM	Plasmodium falciparum 3D7		
REFERENCE	Plasmodium falciparum 3D7		
AUTHORS	Eukaryote: Alveolata; Apicomplexa; Haemosporida; Plasmodium.		
	1 (bases 1 to 561)		
	Tang, R., Cole, R., Chakrabarti, D., Haywood, R., Clifton, S., Pape, D.,		
	Matta, M., Hillier, L., Martin, J., Wylie, T., Dante, M., Theising, B.,		
	Bowers, Y., Gibbons, M., Ritzer, E., Bennett, J., Jentes, E., Ronko, I.,		
	Tsangareishvili, R., Belaygorod, L., Franklin, C., Carr, L., Grow, A.,		
	Magnute, L., Ritchey, J., Wadkins, J., Kennedy, S., Levinso, D.,		
	Waterston, R., Wilson, R. and Sibley, D.		
TITLE	Washu Plasmodium EST Project		
JOURNAL	Unpublished (2001)		
COMMENT	Contact: L. David Sibley		
	Washu Plasmodium EST Project		
	Washington University School of Medicine		
	4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA		
	Tel: 314 286 1800		
	Fax: 314 286 1810		
	Email: est@watson.wustl.edu		
	Library was constructed by R. Haywood. DNA sequencing by:		
	Washington University Genome Sequencing Center For information on		
	obtaining a clone please contact: L. David Sibley		
	(sibley@orcim.wustl.edu), Washington University		
	Seq primer: -40UP from Glbco		
	High quality sequence stop: 412.		
FEATURES	location/Qualifiers		
Source	1..561		
	/organism="Plasmodium falciparum 3D7"		
	/mol_type="mRNA"		
	/db_xref="taxon:36329"		
	/dev_stage="gametocyte (stage III-V)"		
	/lab_host="DH10B (Genesig, Invitrogen, Inc.)"		
	/clone_lib="Plasmodium falciparum 3D7 gametocyte cDNA library"		
	/note="vector: pluscript SK plus; Site_1: EcoRI; Site_2: XhoI; The library was constructed by R Haywood. cDNAs were synthesized from gametocyte poly(A)+ RNA by oligo d(T) priming, size-selected and directionally cloned into the EcoRI (5' end) to XhoI (3' end) sites of the Uni-ZAP XR lambda vector (Stratagene). The primary library was mass excised as phagemid using the Exsist helper phage (Stratagene). Clones were mass excised using the Exsist helper phage (Stratagene), the phagemids were precipitated		

with pEG 8000 and extracted with phenol/chloroform.
Phagemid DNA was electroporated into DH10B cells. Clone
Availability: David Sibley, Washington University."

ORIGIN

Query Match 5.4%; Score 54.2; DB 6; Length 561;

Best Local Similarity 45.0%; Pred. No. 0.084; Indels 0; Gaps 0;

Matches 203; Conservative 0; Mismatches 248; Indels 0; Gaps 0;

179 GTATTCATATGATCTAACTTCTGATGATGATTAATGAAATGAGAGCTGCTGGA 238
15 GTGCTCAATGTTGTTCTTCAAAAAGATCTATTATGTAATATGAGTACTTTGG 74
239 GAATTTATCAAGATATATGCTTAATCTTTCATCTGCACTATTAGCTGAAAATGCTATC 298
75 GTTATATCAAGATATATTAAGATATATATACAGAAATCTGATGATCAACTCATTA 134
299 AAATTTCAAGATATGCGAAATGTTTATTAAGAAAGAACTGATTTGAAAGTCAAT 358
135 AAAAAGGAATCAATATCAAAATTAATATACAAATGAGAAATGATCATCACTAATA 194
359 ATCTCAAAAGATTTGGAATGTTCTGAGCTTGAATAAATCAATGAGTCTATTCAG 418
195 ATATATATATATATATATATATATATGATGAAATTAACAAAGAGTATTTGCTAT 254
419 ATGAGATTAATATATGCACTATATTTTCTCTGACAGAAATTCACCTGAGGAAA 478
255 ATTAATATATATATGTAAGATTTATCTTATGCTTAAATTTAGCTGACGACATGATTA 314
479 ATCAACATCAAAATCCGCAAGATTTTAAATTAATGATTTCTTATTTACCTATCTG 538
315 CTAATACAACTTCAAAAGATTAATTAATGATTAAGTCAAAATTTGAGATTAATAT 374
539 CTGTAATCTCACTGGAAGAGATTTTTCAAAACTTTTACATGATGATTAAGGCTA 598
375 CTGATGTTTCAATTTTAAAGAGCTGGAAGTACATATATTTTACCGTACACAAATTA 434
599 AATCATTAAGATTAATTAATGAGAAAAA 629
435 TAAATATATGATATGATATGATTAATTA 465

RESULT 12

CNS016HF/c 1101 bp DNA linear GSS 26-JUL-1999

LOCUS Drosophila melanogaster genome survey sequence T7 end of BAC

DEFINITION BACN16H14 of DrosBAC library from Drosophila melanogaster (fruit

fly), genomic survey sequence.

ACCESSION AL106749

VERSION AL106749.1 GI:5623618

KEYWORDS GSS.

SOURCE Drosophila melanogaster (fruit fly)

ORGANISM Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 1101)

Genoscope.

Direct Submission

Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :

BP 191 91006 EVRY cedex - FRANCE (E-mail : seque@genoscope.cns.fr

- Web : www.genoscope.cns.fr)

Determination of this BAC-end sequence was carried out as part of a

collaboration with the European Drosophila Genome Project (EDGP) -

http://www.edgp.ebi.ac.uk - This Drosophila melanogaster BAC

library (Dros BAC) was made by Alain Billand at CEPH (Centre

d'Etude du Polymorphisme Humain) with funding provided by a MRC

project grant. The DNA was prepared from embryos by Alain Bucheton

and Genevieve Payan. It has been constructed in the vector

pBelobAC11.

Location/Qualifiers

1..1101

/organism="Drosophila melanogaster"

/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACN16H14"
/clone_1lb="DrosBAC"
/plasmid="pBelobAC11"
/note="end : 17"

ORIGIN

Query Match 5.3%; Score 53.8; DB 9; Length 1101;

Best Local Similarity 27.4%; Pred. No. 0.11;

Matches 150; Conservative 136; Mismatches 259; Indels 2; Gaps 1;

421 GACGATTAATATTAAGCACTATATTTTCTCTGACGGAATTCCTGAGAGAAAT 480
1101 KADDDARRTAAARAAAGARKWMAADTATSWKATWTDVNSAABDRVSVTT 1042
481 CAACATCAAAATGCGCAAGATTTTAAATTAATGATTTCTTATTAACCTTACCT 540
1041 TTWTTTAAKKKKKKDDRDRTDGGKTKTTTTTSAARTSBTTTBTBBSITTTT 982
541 GTAACCTTCACTGGAAGAGATTTTTCAAAACTTTTCAATGATTAAGGCTAAA 600
981 NTWTATAGARBDARRRRDWMKATAAAGGRRBAWRATWMAABAKWMAAAMWMD 922
601 TCATTAGCAATTTATTTGAGAGAAAAAATCTTTCTAACTTTCTTTCGACCCGAG 660
921 TTWTAARRRRGAAGWATTTATRRGAAAAAARAAAGAAATTAATTAATTAAT 862
661 AGATTACCTGAGGAGAAATGTTATTTGCTGAGCAACAGAGAGCGCTTAATGAGA 720
861 VVSSDAATTAATTAASBSAAW--WTKTKTGTTTWTTTTGAATTTGGRAGRGMTGA 804
721 GTGAGTTTAAAGAACTTAAATTAACAAATCTAGAAATGATTTCTAATATGAAAGG 780
803 KDTTATWTGSGGRMTTATGTAATGATGATGATGATGATGATGATGATGATGAT 744
781 GCTGCAAAACAAAGTATGATTCATTTTAAAGAGTCAAAAGGATGATGATGAT 840
743 WAMRAAAMRAAARAAARAAAGAAATTAATTAATTAATTAATTAATTAATTA 684
841 AAGAGCGAAAGATTTGTTACAGCCAGTGCAGTAACTGGAATAATTTGCCGAATAT 900
683 ASSSTTAADGATGKGAANKGDTTARAGTTTAATGGGAAAAATTTGTAATAT 624
901 TTATATAGTGTGAGGCTAAGCAAAAGAGAGGATCTTACTCAAAATGATATGAC 960
623 TATTAMRRRTAARTRAAATVAAAAAATATTAATTAATTAATTAATTAATTA 564
961 AATACAA 967
563 ATTAATA 557

RESULT 13

BX398967

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BX398967 924 bp mRNA linear EST 28-APR-2004
BX398967 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
clone CS0D1067YG04 3-PRIME, mRNA sequence.
BX398967
EST.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 924)
Li, W.B., Gruber, C., Jessee, J., and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
On May 13, 2003 this sequence version replaced gi:30613827.
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France

Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by lile technologies, a
division of Invitrogen. This sequence belongs to sequence cluster
7076.r

For more information about this cluster, see
http://www.genoscope.cns.fr/cdna?c=CS001067BD02NP1&c=7076.r.

FEATURES

SOURCE

1..924
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS001067YG04"
/issue_type="PLACENTA COT 25-NORMALIZED"
/clone_1lb="Homo sapiens PLACENTA COT 25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Query Match 5.3%; Score 53.6; DB 5; Length 924;
Best Local Similarity 37.6%; Pred. No. 0.12;
Matches 158; Conservative 44; Mismatches 218; Indels 0; Gaps 0;

OY 586 GCATTAGAGGCTAAATCTTACGAGATTAATTTGAGAGAAAACCTTCTTAACCTTTC 645
DB 19 GAATTAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAG 78
OY 646 TTGGACACCGCAGAGATTAAGTATGAGAGAGATTAATTTGGCTGAGCAACAGAA 705
DB 79 GGAATGCGGCGCTCTTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 138
OY 706 GCGCCTTAATGAGAGAGATTTTAAAGACTTAATAAATAAATAAATAAATAAATAAATAA 765
DB 139 AAAAAAGAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 198
OY 766 TCTAATATGAGAGGCTGCAAAACAAAGTAAGTATTAATTAATAAAGAGTACAAAG 825
DB 199 CAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 258
OY 826 GGTACGCTCCACAGACGCGAAGTATGCTACAGCAGAGGAGTAACTGTGAA 885
DB 259 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 318
OY 886 AAATGCGCAATTAATTTATAGTGTGAGGCTAACCAAAAAGAGGGTAACTTTTACT 945
DB 319 AAAAAAGAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 378
OY 946 CAAAAATGATGACAAATACAAATGACGGTTCATAGTGTGAACTCATTAATAAATAATA 1005
DB 379 MAATAAGGAGAAATAAAGAGAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 438

RESULT 14

CNS020K7

LOCUS 1092 bp DNA linear GSS 01-SEP-2000
DEFINITION Tetracodon nigroviridis genome survey sequence T7 end of clone
222L11 of library G from Tetracodon nigroviridis, genomic survey
sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

ALI75696
ALI75696.1 GI:7813753
GSS; genome survey sequence.
Tetracodon nigroviridis
Tetracodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Perciformes; Tetraodontiformes;
Tetraodontidae; Tetraodontidae; Tetraodon.
Roest Crolius,H., Jallion,O., Dasilva,C., Bouneau,L., Fisher,C.,

Bernot,A., Fizames,C., Wincker,P., Brothier,P., Quetier,F.,
Saurin,W. and Weissbach,J.
Estimate of human gene number provided by genome-wide analysis
using Tetracodon nigroviridis DNA sequence
Nat. Genet. 25 (2), 235-238 (2000)
10835645

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

JOURNAL

PUBMED

JOURNAL

PUBMED

JOURNAL

PUBMED

JOURNAL

PUBMED

JOURNAL

PUBMED

JOURNAL

PUBMED

JOURNAL

PUBMED

JOURNAL

PUBMED

JOURNAL

PUBMED

JOURNAL

PUBMED

JOURNAL

PUBMED

JOURNAL

PUBMED

JOURNAL

PUBMED

JOURNAL

PUBMED

JOURNAL

PUBMED

JOURNAL

PUBMED

JOURNAL

PUBMED

JOURNAL

PUBMED

JOURNAL

PUBMED

JOURNAL

PUBMED

JOURNAL

PUBMED

JOURNAL

PUBMED

JOURNAL

PUBMED

JOURNAL

PUBMED

JOURNAL

PUBMED

JOURNAL

PUBMED

JOURNAL

PUBMED

JOURNAL

PUBMED

JOURNAL

PUBMED

JOURNAL

REFERENCE
Acanthomorpha; Acanthopterygii; Percomorpha; Gasterosteiformes;
Gasterosteidae; Gasterosteus.
1 (bases 1 to 1327)
AUTHORS
Kingsley, D., Grimwood, J., Dickson, M., Schmutz, J. and Myers, R.M.
TITLE
Expressed sequence tags from Gasterosteus aculeatus
JOURNAL
Unpublished (2004)
COMMENT
Contact: Grimwood, Jane
Stanford Human Genome Center
Stanford University School of Medicine
975 S California Avenue, Palo Alto, CA 94304, USA
Tel: 650 320 5917
Fax: 650 320 5801
Email: jane@hgsc.stanford.edu
Plate: 79
Class: BAC ends
High quality sequence start: 26
High quality sequence stop: 632.
Location/Qualifiers
1..1327

FEATURES
source
1..1327
/organism="Gasterosteus aculeatus"
/mol_type="genomic DNA"
/strain="Salmon River"
/db_xref="taxon:69293"
/clone="CH213-79D13"
/sex="Mixed"
/cell_type="Blood"
/clone_lib="CH213"
/note="Vector: pTABAC2.1; Site 1: EcoRI; The sequence of the clone was established as a mapping and sequencing collaboration at the Stanford Genome Evolution Center, funded by the NIH Centers of Excellence in Genomic Science (CEGS) initiative (<http://cegs.stanford.edu>). The clone was isolated from the BAC library CHORI-213 built by Pieter deJong in collaboration with the Stanford Genome Evolution Center (<http://www.chori.org/bacpac/>). Clones may be purchased from BACPAC Resources (http://www.chori.org/bacpac/ordering_information.htm)."

ORIGIN

Query Match 5.2%; Score 52.4; DB 9; Length 1327;
Best Local Similarity 46.0%; Pred. No. 0.24;
Matches 230; Conservative 0; Mismatches 266; Indels 4; Gaps 2;
149 TAAATTTAGCTTACGCTTGGCTGTAAGTGATCATATATATCTTAACCTTCCGATG 208
776 TATTTTAAATAAATTTAAATTTAATATATATATATATATATATATATATATATAT 835
209 ACTATATATAGATTAAGAGACTGCTGAGAGATTTATCAAGATATATATGCTAATCTT 268
836 AATTTATATTTAAATTTTAAATTTTATTTAATTTTATTTAATTTTATTTTATTA 895
269 CATCTGCACTATTTAGTGATAATGATCAATTTCTAAGATATGCAATGCTTTT 328
896 TTTTAAATTAATTAATAATA--TATTTAATTTTAAATTTTAAATTTTAAATTTTAA 952
329 ATTAAGATGAAGCTGATTTTGAAGTCAATATCTCAAAACATTTGGAATGTTCTGAC 388
953 TTTAAATAATTAATTAATTAATTTAATTTTAAATTTTAAATTTTAAATTTTAAATA 1011
389 TTGAATAATTAACCATTTAGTGTATTTGATGAGATTAATTAATTTAGACATATTTT 448
1012 TTTTAAATTAATTAATTTAATTTAATTTAATTTAATTTAATTTAATTTTAAATA 1071
449 TCTCTGTAAGGAAATTCACCTGAGGAAATCAACAATCAATGCGCAGATTTTAA 508
1072 ATTTATTTAATTTAATTAATTTAATTAATTAATTTAATTTAATTTAATTTAATA 1131
509 AATTATATGATTTCTTATTTACCTTATCTGCTGTAATCTTCACTGGAGAGAGATTTT 568
1132 AAATTAATTAATTTAATTTAATTTAATTTTAAATTTTAAATTTTAAATTTTAAATTT 1191
569 CAAAAACCTTTACATGATGATTAAGGCGTAATCAATTAGAAATTAATTTAGAGAAAA 628

DB 1192 TTTTAAATTTAATTTAATTTAATTTTAAATTAATTTAATTTAATTTAATTTAAT 1251
QY 629 AACTTTCTAAACCTTTCTT 648
DB 1252 ATATATAAATATTTTATTTT 1271

Search completed: November 21, 2004, 08:32:25
Job time : 318 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_n2p model

Run on: November 21, 2004, 08:32:30 ; Search time 139.5 Seconds
(without alignments)
8315.092 Million cell updates/sec

Title: US-10-617-962-2
1335
Perfect score: 1 atggttatacaattacacc.....ctcatataaataatata 1008
Sequence:

Scoring table:
BLOSUM62
Xgapop 10.0, Xgapext 0.5
Xgapop 10.0, Xgapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 1825181 segs, 575374646 residues

Total number of hits satisfying chosen parameters: 3650362

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p_model -DEV=xlp
-O=/cgn2_1/USPTO_spool_p/US10617962/runat_17112004_163922_9749/app_query.fasta_1.1159
-DB=UniProt_02 -QFMT=fastan -SUFFIX=rup -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MTLEN=0 -MAXLEN=200000000
-USER=US10617962@cgn_1_1_152@runat_17112004_163922_9749 -NCPU=6 -ICPU=3
-NO MMAP -LARGESQUERY -NEG SCORES=0 -WAIT -DSPELOCK=100 -LONGLOG
-DEV TIMEOUT=10 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : UniProt_02: *
1: UniProt_sprot: *
2: UniProt_tramb1: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1716	98.9	335	Q7N4K5	Q7N4K5 photorhabd
2	124.5	7.2	1489	Q7RJ32	Q7RJ32 plasmodium
3	119.5	6.9	1485	Q7RCB8	Q7RCB8 plasmodium
4	116.5	6.7	487	Q8ESK1	Q8ESK1 oceanobacil
5	115.5	6.7	473	1	TIG PROMP
6	113.5	6.5	434	Q8I226	Q8I226 plasmodium
7	113	6.5	682	Q6F1M0	Q6F1M0 mesoplasma
8	113	6.5	908	Q8I5F7	Q8I5F7 plasmodium
9	112.5	6.5	1822	Q9U0N4	Q9U0N4 plasmodium
10	111	6.4	995	Q7RT20	Q7RT20 plasmodium
11	111	6.4	1794	2	Q8IHV6
12	110	6.3	3468	2	Q8I104
13	109.5	6.3	414	1	YQ28 BORBU
14	109.5	6.3	1491	2	Q7RN55
15	109	6.3	1119	2	Q7RLZ8
16	108.5	6.3	730	2	Q86A28

17	108.5	6.3	795	2	Q71X29	Q71X29 listeria mo
18	108.5	6.3	795	2	AAT05136	AAT05136 listeria
19	108.5	6.3	900	2	Q8IKV2	Q8IKV2 plasmodium
20	108.5	6.3	1951	2	Q8ILV5	Q8ILV5 plasmodium
21	108.5	6.3	2393	2	Q7RCR2	Q7RCR2 plasmodium
22	108.5	6.3	3482	2	Q8ID46	Q8ID46 plasmodium
23	108	6.2	447	2	Q6LX59	Q6LX59 methanococ
24	108	6.2	447	2	CAF31049	CAF31049 methanoco
25	108	6.2	940	2	Q8Y4N9	Q8Y4N9 listeria mo
26	108	6.2	1127	2	Q9YVY6	Q9YVY6 melanoplus
27	108	6.2	2612	2	Q8I5X5	Q8I5X5 plasmodium
28	107.5	6.2	1092	2	Q6F029	Q6F029 candida glia
29	107.5	6.2	1171	2	Q8I3F4	Q8I3F4 plasmodium
30	107.5	6.2	2558	2	Q8IKR6	Q8IKR6 plasmodium
31	106.5	6.1	499	2	Q6U7S2	Q6U7S2 crinipellis
32	106.5	6.1	499	2	AAQ74285	AAQ74285 crinipell
33	106.5	6.1	867	2	Q7REJ4	Q7REJ4 plasmodium
34	106	6.1	353	2	Q8I7S1	Q8I7S1 plasmodium
35	106	6.1	412	2	Q7RT80	Q7RT80 plasmodium
36	106	6.1	1330	2	Q8IBJ3	Q8IBJ3 plasmodium
37	106	6.1	1636	2	Q8ID77	Q8ID77 plasmodium
38	106	6.1	1785	2	Q25685	Q25685 plasmodium
39	106	6.1	2740	2	Q7RES2	Q7RES2 plasmodium
40	105.5	6.1	460	2	Q89627	Q89627 clostridium
41	105.5	6.1	512	2	Q7RKX6	Q7RKX6 plasmodium
42	105.5	6.1	744	2	Q7RS98	Q7RS98 plasmodium
43	105.5	6.1	750	2	Q6MSZ0	Q6MSZ0 mycoplasma
44	105.5	6.1	750	2	CAE77248	CAE77248 mycoplasma
45	105.5	6.1	755	2	Q56271	Q56271 human herpe

ALIGNMENTS

RESULT 1

Q7N4K5 ID Q7N4K5 PRELIMINARY; PRT; 335 AA.
AC Q7N4K5;
DT 01-MAR-2004 (TREMBLrel. 26, Created)
DT 01-MAR-2004 (TREMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
DE Unknown protein.
GN OrderedLocusNames=plu2326;
OS Photorhabdus luminescens (subsp. laumondii).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Photorhabdus.
OX NCBI_TaxID=141679;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=T01;
RX MEDLINE=22957627; PubMed=14528314;
RA Duchaud E., Rusniok C., Frangeul L., Buchrieser C., Givaudan A.,
RA Taourit S., Bocs S., Boursaux-Eude C., Chaudier M., Charles J.-F.,
RA Daes B., Derose R., Derzelle S., Freysinet G., Gaudreau S.,
RA Medigue C., Lanois A., Powell K., Sigler P., Vincent R., Wingate V.,
RA Zouine M., Glaser P., Boemare N., Danchin A., Kunst F.,
RT "The genome sequence of the entomopathogenic bacterium Photorhabdus
RT luminescens".
RL Nat. Biotechnol. 21:1307-1313(2003).
DR EMBL; BX571866; CAE14619.1; -.
KW Complete proteome.
SQ SEQUENCE 335 AA; 38038 MW; ASE7P93A73EA50B0 CRC64;

Alignment Scores:

Pred. No.: 1,26e-125 Length: 335
Score: 1716.00 Matches: 331
Percent Similarity: 99.10% Conservative: 1
Best Local Similarity: 98.81% Mismatches: 3
Query Match: 98.90% Indels: 0
DB: 2 Gaps: 0

US-10-617-962-2 (1-1008) x Q7N4K5 (1-335)

OY	I	ATGCTTATACATTAAACCCTGATGATAGAAGGAGTATTCACCCTGTGA AAAACAATA	60
Dp	1	MetValIleGlnLeuThrProAspSerGlyTyrProProValGluylsGlnIle	20
OY	61	GCAGAGATATATGTA CGTATAC TTA AACTTTAAGCAAAACGATAGAGGGTGCTAACGATCA	120
Dp	21	AlaGlyAspIleValArgIleLeuasnPheylsGlnThrAspGluGlyHisTrpAlaSer	40
OY	121	TATGGAAATGGAATATCAGCCTAAGAAAATAATATATTAGCTTACCGCTTGGCTGAAGTGT	180
Dp	41	TyrgIylIeGluTyfArgAlaIysLysIleIleLeuAlaTyfAlaLeuAlaValSerGly	60
OY	181	ATTGCTAATGTA TC TTA AACTTCCTCATGACTATTAATTAAGAAATTAAGAGACTGCTGAGAGA	240
Dp	61	IleHisasnValSerLysLeuProAspPtyrTfLysasnLysglutThrAlaGluArg	80
OY	241	ATTATTCAGAGATTAATGTCCTAATCCTTATCTGATCTGCACTATTAGTGA AAAATGSGTATCAA	300
Dp	81	IleTyrgInglutyfMetSerasnLeuSerSerAlaLeuLeuGluylsGluasnGlyAspGln	100
OY	301	ATTTC TTA AAGATATGAGCAAAATG GTTTT TA AAGAAATGA ACTG GATTTT TGA AGGCTCAATAT	360
Dp	101	IleSerLysAspMetAlaasnGlyPheTyfLysasnGluLeuasnPhelGluGlyIntyr	120
OY	361	CCTCAAAACATTTGGAAATGTTCTCTGAGCTTGAAAAATPA AACATTGAGTGTCTTATTCAGAT	420
Dp	121	ProGlnasnlIeTyfAsnValProGluLeuGluasnLysProLeuSerAlaTyrSerAsp	140
OY	421	GACGATTAATTA TT TGACATATATTTTTCTCTGTGACAGAAATTCACATCGAGGAGAAAT	480
Dp	141	AspAspLysLeuLeuAlaLeuTyfPhePheSerValGlnGluIleProLeuGluGluAsn	160
OY	481	CAACATCAAAATGCCGCAAGATTTT TTA ATTAATGATTTCTTATTTACCTTATCTGTCT	540
Dp	161	GlnGlnSerAsnAlaAlaArgPhePheLysIleuIleAspPheLeuThrLeuSerAla	180
OY	541	GTAACCTTCACTGGGAAAGAGATTTTTC AAAAACTTTTACATGAGATTAGAGCTAAA	600
Dp	181	ValThrSerLeuGlylAArgArgIlePheSerLysAsnPhetyfzAnsglyLeuGlnSerLys	200
OY	601	TCATTAGGAATTAATTA TTGAGGAAAAAACTTCTTAACCTTCTTGGACACCGCCAG	660
Dp	201	SerLeuGlnAsnTyfIleGluIuArgLysLysPheProLysPheProPheArgProproGln	220
OY	661	AGATTACCTGATGGCAGATAGATTATTTTGCTGCACCAACAGACG CCTTAATGGAGA	720
Dp	221	ArgLeuProAspGlylAArgIleGlyTyfLeuAlaGlyProThrGluAlaPhePolystyParg	240
OY	721	GTGAGTTTAAAGAACTTAA AAATAATCAATCTGAGATGAGATTTTCTAATATGAGAGG	780
Dp	241	ValSerPheLysgluLeuLysasnAsnLysSerArgAsnGlyPheSerAsnMetGluGly	260
OY	781	GCTGGA AAACAAGATAGTTCATTTATTA AAAGAGGTACAAAAGGTAA CGCTCCACAG	840
Dp	261	AlaAlaLysGlnLysTyfSerSerPheIleLysGlnValGlnLysglYasnAlaPheGln	280
OY	841	ACAGCAGCGAAAAGTATGTCACGACGACGAGTCAGTACCTGGA AAAAATGCCGAATAT	900
Dp	281	ThraAlaAlaLysSerIleGlyThrAlaSerGlySerAsnLeuGluLysLeuProAsnAsn	300
OY	901	TTATATAGTGTAGGCTTAAGCCAAAAGA CAGGGTAACTTTA CTCAAAATGATACTGAC	960
Dp	301	LeuTyfSerValArgLeuSerGlnLysAspArgValThrPheThrGlnAsnAspThrAsp	320
OY	961	AATACAAATGACGGTTCATATGCTTGGAACCTCATTAATTA AAATXTA 1005	
Dp	321	AsnThrMetThrValHisSerValGlyThrHisTyfLysAsnIle 335	
RESULT 2			
ID	Q7RJ32	PRELIMINARY;	PRT; 1489 AA.
NC	Q7RJ32;		
DT	01-MAR-2004	(TRENBLREL. 26, Created)	

Dt	01-MAR-2004 (TrEMBLrel. 26, last sequence update)
Dt	01-MAR-2004 (TrEMBLrel. 26, last annotation update)
Ds	Rhodtry protein-related (Fragment).
Gn	Pname=PY03432;
Oc	Plasmodium yoeiii Yoeiii.
Ox	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
Rn	(1)
Rp	SEQUENCE FROM N.A.
Rc	STRAIN=17XNL;
Rx	PubMed=13368865;
Ra	Carlton J.M., Angiuoli S.V., Suh B.B., Kooij T.W., Pertea M.,
Ra	Silva J.C., Ermolaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,
Ra	Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Biwell S.L.,
Ra	Shallom S.J., van Aken S.E., Riedmiller S.B., Feldblum T.V.,
Ra	Cho J.K., Quackenbush J., Sedegah M., Shoabdi A., Cummings L.M.,
Ra	Florens L., Yates F.R., III, Rahne J.D., Sinden R.E., Harris M.A.,
Ra	Vannitham D.A., Pfeister P.R., Bergman L.W., Vaidya A.B.,
Ra	van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,
Ra	Satzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,
Rt	Carucci D.U.;
Rt	"Genome sequence and comparative analysis of the model rodent malaria
Rt	parasite Plasmodium yoelii yoelii.";
Rl	Nature 419:512-519(2002).
Cc	-1- CUTION: The sequence shown here is derived from an
Cc	EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
Cc	preliminary data.
Dc	EMBL; AABL01000986; EAA23001.1; -
Dc	InterPro; IPR011591; Botulinum.
Dc	InterPro; IPR006499; Reticulocytebp.
Dc	Prodont; PD001963; Botulinum; 1.
Dc	TIGRfams; TIGR01612; 23SKda-fam; 1.
Ft	NON TER 1 1
SO	SEQUENCE 1489 AA; 173699 MW; 512P2A9A383D8A3F7 CRC64;
 Alignment Scores:	
Pred. No.:	0.346 Length: 1489
Score:	124.50 Matches: 66
Percent Similarity:	39.328 Conservative: 61
Best Local Similarity:	20.43% Mismatch: 153
Query Match:	7.18% Indels: 43
DB:	2 Gaps: 12
 US-10-617-962-2 (1-1008) x Q7RJ32 (1-1489)	
Oy	136 CGAGCTAGAATAATATTACGCTTTCGCTGAAT----- 1777
Dd	393 LySserAsnlyeAlalEalaIleLySlSerlleAepPropheylThrySpHe 4122
Oy	178 ---GGATTCAATATGTACTCTTAACCTCCATGACATCTATTAAAGATAAAGACGCT 2344
Dd	413 LeuylleLennaspIlleLySlySlySerAspGlnTyLeuLyGluThrglaSenlle 4322
Oy	235 GAGAGAAATTTACAAGATATATGCTATCTTTCA-----TCGCACATTTA 2822
Dd	433 GluLys-----GluLeSerLennuSerlleAephThrgInguThryLeu 4488
Oy	283 GGTTGAAAATGCGTCAAAATTTCT-----AAAGTATGCGAAATGTTTTATAGCAT 3366
Dd	449 LyggluAsngclYAspyLeuSerThrLeuLySgluLeuGluSerLeuLySaEngln 4688
Oy	337 GAACGTGATTTGAAGCTGCAATATCTCATAAACATTGGATGTCTCGAGCTTGAAAT 3966
Dd	469 LySlyAsenllleGlunspTygLySlysgLu-----LeuAspGluValAsenSer 4844
Oy	397 AACCATTTGAGTGCCTTATTCAGATGACATGACATTAATTTATTTCTCTGTA 4566
Dd	485 LyatlleGuLuglnlleGuLyrghApAlaEnglnSerLySlySaenTyrgluilleGlyle 5044
Oy	457 CAGGAAATTCCTGAGGAGAAAAATCAACAATCAAATGCCGACAGATTTTAAATTAATT 5166
Dd	505 ValGlu---LySlleLennsluileaglulaSenlySlySAgllleGluSerThryls 5233

QY 517 GATTTCCTATTT---ACCTATCTGTCTAATCTCACTGGAAGAGGATTTTTCANAA 573
 Db 524 GLeuLeuIeGlnProthrlIeGlnAsnLeuIeSerThrGlnGluLeuLeu---Lys 542
 QY 574 AACTTTTCAATGGATTGAGGCTTAATGATTAAGATTTATTTAGAGAGAAAACTT 633
 Db 543 AsnIleGlnHsIglAsnLysAlaLysSerTyrLeuAspTyrValLysGluAspGluPhe 562
 QY 634 TCTAAACCTTTCTTTCAGCACACCGAGATTAAGTACGAGCAATG----- 681
 Db 563 AspArgIleValIthrHisLeuLysLysLeuAspAsnValAsnValLysPheThrAsn 582
 QY 682 GGTATTGTCGTGACCAACAGAGCGCTTAATGAGAGTGAATTTTAAAGAACTTAA 741
 Db 583 GluTyrLeuLysAlaAsnGluGlyPheAspAsnIleSerLysSerIleAsnAsnValLys 602
 QY 742 AATAAATAATGATGATGATTTTCTAATATGAGAGGCGCTCAAAAATAATATAGT 801
 Db 603 AsnSerThrAspGluAsnSerLeuLeuAsnIleLeuAsnGlnThrLysGlnMetTyrGlu 622
 QY 802 TCATTATTAAGAAGAGTACAAAAGGT----- 831
 Db 623 AsnIleValSerLysThrTyrAsnSerTyrLysTyrAspAlaGluAsnIlePheIleAsn 642
 QY 832 GCTCCACAGACAGACGCGAAA-----AGTATTGTCAGACGAGTGCAGTAACTG 882
 Db 643 IleProLysLeuAlaAsnSerLeuAsnIleGlnIleLysSerSerGlyIleAspLeu 662
 QY 883 GAAAATATGCGCAAT-----AATTATATAGTGTAGAGGCTTAAGCAAAAAGACGGTA 936
 Db 663 PheLysAsnIleAsnIleAlaIleLeuProTyrLeuAspSerGlnLysAspThrLeu 682
 QY 937 ACCTTACTCAAAATGATGATGACATGACATGACGTTCAAGTTCGTAATCATATAT 996
 Db 683 ThrPheIleProSerProGlnLysThrSerGlnThrLysIleSerAspSerTyr 702
 QY 997 AAAAATATA 1005
 Db 703 AsnThrLeu 705
 RESULT 3
 QYRCB8 PRELIMINARY; PRT; 1485 AA.
 AC QYRCB8;
 DT 01-MAR-2004 (TREMBLrel. 26, Created)
 DT 01-MAR-2004 (TREMBLrel. 26, Last sequence update)
 DE RNA recognition motif, putative.
 GN Name=PY05866;
 OS Plasmodium yoelii yoelii.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=73339;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=17XNL;
 RX PubMed=12368665;
 RA Carlton J.M., Angiuoli S.V., Suh B.B., Kocij T.W., Petrea M.,
 RA Silva J.C., Emtolaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,
 RA Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,
 RA Shallom S.J., van Aken S.B., Riedmuller S.B., Feldblum T.V.,
 RA Cho J.K., Quackenbush J., Sedegah M., Shoaiti A., Cummings L.M.,
 RA Florens L., Yates F.R., III, Raine J.D., Sinden R.E., Harris M.A.,
 RA Cunningham D.A., Preiser P.R., Bergman L.W., Valdya A.B.,
 RA van Lin L.H., Jansz C.J., Waters A.P., Smith H.O., White O.R.,
 RA Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,
 RA Carucci D.J.;
 RA "Genome sequence and comparative analysis of the model rodent malaria
 RT parasite Plasmodium yoelii yoelii.";
 CC Nature 419:512-519 (2002).
 CC -I- CAUTION: The sequence shown here is derived from an
 CC EMBL/GenBank/DBD whole genome shotgun (WGS) entry which is
 CC preliminary data.
 DR EMBL, AABL01001921; EAA17961.1; -.

DR InterPro; IPR000504; RNA_rec_mot.
 DR Pfam; PF00076; RRM 1; 1.
 DR PROSITE; PS50102; RRM; 1.
 DR PROSITE; PS00030; RRM RNP 1; UNKNOWN 1.
 SQ SEQUENCE 1485 AA; 172775 MW; E12120E6F2E1CB CRC64;
 Alignment Scores:
 Pred. No.: 0.852 Length: 1485
 Score: 119.50 Matches: 79
 Percent Similarity: 35.97% Conservative: 62
 Best Local Similarity: 20.15% Mismatches: 126
 Query Match: 6.89% Indels: 125
 DB: 2 Gaps: 19
 US-10-617-962-2 (1-1008) x QYRCB8 (1-1485)
 QY 22 GATGATGAGAGTGAATATCCACCGCTTGAAGAAAGCAATATGACGAGATATGATGATA 81
 Db 37 AsnAspAsnGlnGlnPheGlnGluSerArgLysTyrIleHisArgGluIleThrArgThr 56
 QY 82 CTAAAC-----TTTAAGCAACAGATGAGGTCATACGATCATAT 123
 Db 57 LeuAsnAlaAsnSerMetLysSerTyrArgAsnSerGlnGluSerHisLeuGlnAsn--- 75
 QY 124 GGAATTGAAATATCGAGCTTAAGAAATATATATTAAGCTTAAGCTTGGCTGTAAGTAT 183
 Db 76 SerGlyHisTyrLys-----TyrAlaSerThrAsnAlaGluVal 88
 QY 184 CATATGATGATTAATCTTCTGATGATGATTAATTAAGATTAAGAGACTGTCAGAGAAAT 243
 Db 89 GlnAsnLysLysGlnPheSerAsnAsn-----AsnLysAspAspValGluLeuLys 105
 QY 244 TATGAGAAATATATGCTTAATCTTCAATCTGCACTTAAGTGAATAATGTCATCAAT 303
 Db 106 ArgProLysAsnTyrAsnAsnTyrGlyAsnTyrLeuAsnThrAsnTyrAsnAsnGlyIle 125
 QY 304 TCTAAAGATATGCAAAATGTTTATTAAGATGATGATGATGATGATGATGATGATGAT 363
 Db 126 AsnTyrLysAsnValAsnGlnLysLysAsnAsnAsnAlaValLeuSerGlyGlnLysPro 145
 QY 364 CAATAATTTGCAAT-----GTTCTGAGCTTGAAT- 396
 Db 146 HisThrIleGlnAsnGluLeuIleValAsnAspHisLysAsnGlnTyrHisAsnGlnLysArg 165
 QY 397 ---AAACATGAGTGGCTTATTCAGATGACGATTA- 429
 Db 166 LysLysAsnGlnLysSerTyrGluAspLysAsnLysMetValGluThrLeuLysAsnGlu 185
 QY 430 -----TTATTAGCACTATATTTTCTCTGACAGAAATTCACCTGAGGAAAT 480
 Db 186 LysLysAsnLeuLeuLysHisGluPhePheTyrGlyGluSer-----GlnGluAsn 202
 QY 481 CAACATCAAAATGCCGCAAGATTTTAAATTAATGATTTCTATTATTAACCTTA----- 534
 Db 203 LysArgIleAsnGln-----LysIleGluAsnPheIlePheThrAsnGlnGlu 218
 QY 534 ----- 534
 Db 219 LysLysAsnTyrSerThrAsnAsnAsnGlnTyrAsnLysTyrGluValAsnArgAsnAsn 238
 QY 535 ---TCTGCTGAATCTTCACTGGAGAGAGATTTTTCAAAATACTTTTAAATGATGATTA 591
 Db 239 AspAsnIleValIglAsnIleAsnMetAsnValPheAsnHisAsnGlnPheAsnAsnIle 258
 QY 592 GAGGCTAAATCATTAAGAGATTAATTAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 651
 Db 259 LysLysLysAsn-----AsnLeuValSerMetLys----- 268
 QY 652 CCACCGCAGAGATTAACCTGATGCGAGATAGATTATTTGGCTGACCAAGAGCGCT 711
 Db 269 -----AspGly----- 270
 QY 712 AATGAGAGTGAATTTTAAAGAACTTAATAAATAATCAATCTAGGAATGA----- 762

```

Db      271 -----||| |||||:|:|:| |||
QY      763 -----TTTTTAATATGAGAGGGCTGCAAAACAAAGTAT-----AGTTCAATT 807
Db      284 HSH1SPheAsnAsnGlnAsp-----IleGluValGluTyrArgAspLysAsnSerTyr 301
QY      808 ATAAAAGAGCTACAAAAGGTTACGCTCCACAGACAGCAGCAAAAGTTTGTTGTCAGCC 867
Db      302 IleAsnAspPhe---LysSerLysIleProGlnTyrAsnSerLysAsnAlaPheSer 320
QY      868 AGTGACGATTAACCTGGAATAATTGCGGAT-----AATTTA 903
Db      321 LysGlySerAsnArgAspIleSerAsnAsnIleAsnAsnAlleGlyValSerAsnMet 340
QY      904 TATATGTGAGGCTAGACCAAAAAGACAGGTTACTTACTCAAAATGATGACAAT 963
Db      341 AsnSerTyrArgMetSerCylAsnAsnMetTyrArgTyr-----AspThrLysAsn 357
QY      964 ACAATGACGTTCAATAGTGTGAACTCATTTATAA 999
Db      358 IleMetGlnIleAsnProLysAsnIleHISGluLys 369

```

RESULT 4

```

Q8ESK1 ID Q8ESK1 PRELIMINARY; PRT; 487 AA.
AC Q8ESK1;
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
DE Catalase (EC 1.11.1.6).
GN OrderedLocusNames=O80633;
OS Oceanobacillus thelyensis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Oceanobacillus.
OX NCBI_TaxID=182710;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HTE831;
RX MEDLINE=22220767; PubMed=12235376;
RA Takami H., Takaki Y., Uchiyama I.;
RT "Genome sequence of Oceanobacillus thelyensis isolated from the Iheya
RT Ridge and its unexpected adaptive capabilities to extreme
RT environments.";
RL Nucleic Acids Res. 30:3927-3935(2002).
CC -1- FUNCTION: Decomposes hydrogen peroxide into water and oxygen;
CC serves to protect cells from the toxic effects of hydrogen
CC peroxide (by similarity).
CC -1- CATALYTIC ACTIVITY: 2 H(2)O(2) = O(2) + 2 H(2)O.
CC EMBL, AP004595; BAC12589.1; -.
DR HSP; P04040; IDGF.
DR GO; GO:0004096; F:catalase activity; IEA.
DR GO; GO:0016491; F:oxidoreductase activity; IEA.
DR GO; GO:0004601; F:peroxidase activity; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR GO; GO:0006979; P:response to oxidative stress; IEA.
DR InterPro; IPR002226; Catalase.
DR Pfam; PF00199; Catalase; 1.
DR PRINTS; PR00067; CATALASE.
DR ProDom; PDO00510; Catalase; 1.
DR PROSITE; PS00437; CATALASE_1; UNKNOWN_1.
DR PROSITE; PS00438; CATALASE_2; 1.
KW Complete proteome; Heme; Hydrogen peroxide; Iron; Oxidoreductase;
KW Peroxidase.
SQ SEQUENCE 487 AA; 55391 MW; 841B2B318C7ACAB CRC64;

```

Alignment Scores:

```

Pred. No.: 1.38 Length: 487
Score: 116.50 Matches: 72
Percent Similarity: 35.11% Conservative: 40
Best Local Similarity: 22.57% Mismatches: 86
Query Match: 6.71% Indels: 121
DB: 2 Gaps: 18

```

US-10-617-962-2 (1-1008) x Q8ESK1 (1-487)

```

QY      160 TACGCTTTGGCTGTAAAGTGTATTCATATGTAATTAACCTTCGTGATGAC----- 210
Db      168 TrpSerLeuSerProGluSerLeuHISGlnValThrTyrLeuHISerAspArgGlyLeu 187
QY      211 -----TATTATAGAAATTA 225
Db      188 ProAlaThrLeuArgHISMetAsnGlyTyrGlySerHISThrPheLysTyrAlaAsnAla 207
QY      226 GAGACTGCTGAGAGATTTATCAAGATAT-----ATGCTAAT 264
Db      208 Glu---GlyGluAlaPheTyrValLysTyrHISPhelIleSerGluGlnGlyIleLysAsn 226
QY      265 CTTCATCTGCACTA-----TTAGGTGAAAATGGTGATCAAAATTTCTAAAGAT 312
Db      227 LeuAspProAlaLeuAlaAspLysIleAlaGlyGluAsnProAspTyrHISThrGluAsp 246
QY      313 ATGCAAAATGGTTTTTATTAAGATGACAGCTGATTTTGAAGTCAATATCCCAAAACATT 372
Db      247 LeuPheAsnAlaIleGlu-----GluGlyAspHisPro-----Ala 258
QY      373 TGAATGTCTCGAGCTTGAATAATTAACCATTAAGTGCTTATTCAGATGACGATTAATTA 432
Db      259 TrpLys----- 260
QY      433 TTAGCATATATTTTCTCTGTACAGAAATTCACATGAGAAATCAACATCAAAAT 492
Db      261 -----LeuTyr-----ValGlnIleMetProTyrGluAsp----- 270
QY      493 GCCGCAAGATTTTAAATTAATGATTTCTATTATTAACCTATTCGCTGTAACTGACTG 552
Db      271 --AlaLysThrTyrLysTyrAspProPheAspValThr----- 282
QY      553 GGAAGAGAGATTTTTCAAAACCTTTACAATGATGATGAGCT----- 597
Db      283 -----LysIleTyrSerAspAlaHISArgTyrArgLeuGlyAlaAsnHISerIleProVal 300
QY      598 --AAATCATTAAGAAATTAAT-----ATTGAGAGAAAAAACTTTCTAAACCTTTC 645
Db      301 AsparGAsnProGluAsnTyrPheAlaGluValGluGlnAlaIaPheSerProGlyGln 320
QY      646 TTTCGACCA-----CCGACAGATTAACCTGATGCGACAAATAGGTTATTG 690
Db      321 PheValProGlyIleGluAlaSerProAspLysMetLeuGlnGlyArg-----Leu 337
QY      691 GCTGACCAACAGAGGCTTAATGAGAGTGAGTTT----- 729
Db      338 PheGlyTyrSerAspAlaHISArgTyrArgLeuGlyAlaAsnHISerIleProVal 357
QY      730 -----AAAGACTTAATAAATTAACAATCTAGAAATGATTTTCTAATATGGAAGG 780
Db      358 AsnArgProLysAsnGluValAsnAsnTyrGlnArgAspGlyPheMetSerValAspGly 377
QY      781 GCTGCAAAACAAAGTATAGTTCAATTATTAAGAAGGTACAAAAGGTTAAGCTTCACAG 840
Db      378 -----AsnGlyGlyAsnLysProAsn 384
QY      841 ACAGCAGGAAAGTAT--GTTACAGCCAGTGCAGTAACCTCGAATAAATTTGCGGAT 897
Db      385 TyrGluProAsnSerValAsnGlyProThrGluAspSerGlyAlaLysIleAsnProPhe 404
QY      898 AATTATATATAGTGTAGGCTAAGCAAAAAGACAGGTTACTTACTCAAAATGAT 954
Db      405 GlnIleTyr-----GlyGlnAlaAspSerValThrTyrAspSerAspAsp 419

```

RESULT 5

```

TIG_PROMP ID TIG_PROMP STANDARD; PRT; 473 AA.
AC Q7U2K8;
DT 29-MAR-2004 (Rel. 43, Created)
DT 29-MAR-2004 (Rel. 43, Last sequence update)
DT 01-OCT-2004 (Rel. 45, Last annotation update)

```

DE Trigger factor (TF).
 GN Name=trig; OrderedlocusNames=PMM1655;
 OS Prochlorococcus marinus subsp. pastoris (strain CCMP 1378 / MED4).
 CC Bacteria; Cyanobacteria; Prochlorophytes; Prochlorococcaceae;
 CC Prochlorococcus.
 OK NCBI_taxid=59919;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22825698; PubMed=12917642; DOI=10.1038/nature01947;
 RA Rocap G., Latimer F.W., Lamerdin J.E., Malfatti S., Chain P.,
 RA Ahlgren N.A., Arcilano A., Coleman M., Hauser L., Hess W.R.,
 RA Johnson Z.I., Land M.L., Lindell D., Post A.F., Regala W., Shah M.,
 RA Shaw S.L., Steglich C., Sullivan M.B., Ting C.S., Tolonen A.,
 RA Webb B.A., Zinser E.R., Chisholm S.W.;
 RT "Genome divergence in two Prochlorococcus ecotypes reflects oceanic
 niche differentiation".
 RL Nature 424:1042-1047(2003).
 CC -1- FUNCTION: Involved in protein export. Acts as a chaperone by
 maintaining the newly synthesized protein in an open conformation
 (By similarity).
 CC -1- SIMILARITY: Belongs to the FKBP-type Ppiase family. Trg subfamily.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: BX572094; CAE20114.1; -
 DR HAMAP: MF_00303; -; 1.
 DR InterPro: IPR001179; FKBP_Ppiase.
 DR InterPro: IPR008880; Trigger_C.
 DR InterPro: IPR008881; Trigger_N.
 DR InterPro: IPR005215; Trig_fac.
 DR Pfam: PF00254; FKBP_C; 1.
 DR Pfam: PF05698; Trigger_C; 1.
 DR Pfam: PF05697; Trigger_N; 1.
 DR TIGFams: TIGR0115; Cig; 1.
 DR TIGFams: TIGR0115; Cig; 1.
 DR PROSITE: PS00059; FKBP_PPIASE; 1.
 KW Cell division; Chaperone; Complete proteome; Isomerase; Rotamase.
 FT DOMAIN 174 261 Ppiase, FKBP-type.
 SQ SEQUENCE 473 AA; 53533 MW; 31142AP2CB4E9AED CRC64;
 Alignment Scores:
 Pred. No.: 1.65 Length: 473
 Score: 115.50 Matches: 76
 Percent Similarity: 39.88% Conservatve: 52
 Best Local Similarity: 23.68% Mismatches: 146
 Query Match: 6.66% Indels: 47
 DB: 1 Gaps: 15
 US-10-617-962-2 (1-1008) x TRG_PROMP (1-473)
 QY 43 CCCGTTGAAGCAATAGCA-----GAGATATAGTACTATTAACCTT----- 90
 DB 165 ProValaasnaryproalalyaleuglyalea--ValValserPheylsgly 183
 QY 91 -----AAGCAACAGATGAGGTCATACATCATATGAAATGGA 132
 DB 184 ValTYrlysbapsetelyelsglyleapsglylysetserabsermetekleupen 203
 QY 133 TATCGAGCTAAGAAATATATTAAGCTTACGCTTGGCTGTAAAGTGATTCAT----- 186
 DB 204 LeuglulysasnlymetleleProglyPheValgluglyllyleValglymetlyslleasp 223
 QY 187 -----AAGTATCTAACTCTCTGATGACTAT--TATAAGAAATAAGACT 231
 DB 224 AspaenlysrthleuthrleuargPhepProgluabpTyserHlsGlusperarggly 243
 QY 232 GCTGAGCAATTTATCAAGAATATATGTCTAATCTTTCACTGCACTTTAGTGAAAT 291

DB 244 LysglulalalePheaspIleSerleuylsaspLeuylsGluLeuProgluLeu 263
 QY 292 GGTGATCAAAATTTCTAAAGATATGCAAAATGTTTATTAAGAAATGAACTG-----GAT 345
 DB 264 AsnspasPhealalybglinserclyAsnlysbaserleuylsGluLeuylsAsp 283
 QY 346 TTGAGAGCTCAATATCTCAAAACATTTGAAATGCTTCTGAGCTTGAATAATTAACCAT 405
 DB 284 lIleglulysgluLeuylsGluAsnspheapsanthrgin-----LysasnlyleVal 301
 QY 406 AGTCTTATTAAGAT-----GAGATTAATTTATGACATATATTTTCTCTGTACAG 459
 DB 302 GluAlaLeuMetaspAlaLeuSerlysegluLeuaspAlaGluileProlyalaMetile 321
 QY 460 GAAATTCACCTGAGGAAATCAACAATCAATGATCCGCAAGATTTTAAATTA----- 513
 DB 322 AspllegluValargAsnasnIlegluInthrAlaGlnargPhealaglnGlnlymet 341
 QY 514 --ATGATTTCTTATTAATCTTATCTGCTGTAACCTGCTGGA-----AGAGG 561
 DB 342 AsplleuyserrthPheThrProgluLeuVallyserleuAlaGluSerThrArgPro 361
 QY 562 ATTTTCAAAAACCTTTACAAATGATTAAGGCTTAATGATTAATTAATTTAG 621
 DB 362 GlnAlaGluylsasnValGlnargAsnleuAlaLeuylsAlaLeuSer-----Glu 378
 QY 622 AGAAAAAATCTTTCAAACTTTCTTGACACCGAGATTAATTAATTAATTAATTAATTA 681
 DB 379 ArggluylslethrValaspAsnleuGluileaspGlnlyswelyGlu----- 395
 QY 682 GGTATTTGGCTGAGCAACAAAGAGCGCTTAATGAGATGATGATTTTAAAGACTTAA 741
 DB 396 ---TyrgluaspGluileserlyserProlys---GlnleaspIleGlnlyseuyls 413
 QY 742 AATAACAAATGATGAGATGATTTCTTAATATGGAAGGCGCTGCAAAACAAATATGT 801
 DB 414 ---AspValValArgAsnaspLeuLeuGlnGluilethrTrpleuGluasn 432
 QY 802 TCATTTATAAAGAGCTCAAA-----AAGGTAAAGCTCA 837
 DB 433 SerAlaVallysgluileasngluylsAlaThrlyseuThrlyThrlyThrly 452
 QY 838 CAGACAGCAGGAGAAAGATGATGATGATGATGATGATGATGATGATGATGATGAT 897
 DB 453 AlaThrlyThrlysglyVallyThrlyserlyProlyValAsnlysbgluyls 472
 QY 898 AAT 900
 DB 473 Asn 473
 RESULT 6
 Q81226 PRELIMINARY; PRT; 434 AA.
 ID Q81226;
 AC Q81226;
 DT 01-MAR-2003 (TREMBlrel. 23, Created)
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
 DE TatD-like deoxyribonuclease, putative.
 GN Name=PRF0580C; Synonyms=MALIP3_01;
 OS Plasmodium falciparum (isolate 3D7).
 CC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OK NCBI_Taxid=36329;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22255708; PubMed=12368867;
 RA Hall N., Pain A., Barriman M., Churher C., Harris B., Harris D.,
 RA Mungall K., Bowman S., Atkin R., Baker S., Barron A., Brooks K.,
 RA Buckee C.O., Burrows C., Cherevach I., Chillingworth C.,
 RA Chillingworth T., Christodoulou Z., Clark L., Clark R., Corson C.,
 RA Cronin A., Davies R., Davis P., Dear P., Dearden F., Doggett J.,
 RA Fellwell T., Goble A., Goodhead I., Gwilliam R., Hamlin N., Hance Z.,
 RA Harper D., Hauser H., Hornsby T., Holroyd S., Horrocks P.,
 RA Humphrey S., Jags J.K., James K.D., Johnson D., Kerhornou A.,


```

Db 126 GluYrThrAlaValIleSerAsp-----LysHISglYLeuVal 139
Qy 346 TTGAAGTCATATCCCAAAACATTGGATGTCCTGAGCTGGAATAAATCAATG 405
Db 140 AsnAspLysGlnIYrSerAsnAsn-----HisSerLeuAsnProHe 153
Qy 406 AGTCCTTATTCAGATGACGATAAATTA---TTAGCACTATATTTTCTGTACAGAA 462
Db 154 AsnLeuPheGluAspAspGluIleAsnIYrSerIleYrYr----- 168
Qy 463 ATTCACCTGAGGAATAATCAACATCAATGCCGCAAGATTTTAAATTAATGATTG 522
Db 169 -----LysGluYrSerGluAsnGlnIurHeuThrAgtIprGlnValLeuGluY 186
Qy 523 TTATTTACCTTATTCGCTGATCTTACTGACGGAAGAGAGATTTTCAAAAACTTTTAC 582
Db 187 GlnGluAsnGlnIurPAsnIleProSerIleGluAlaLeuAsnGlnYglYAsnAsnHeuYr 206
Qy 583 -----AATGATTAAGAGGCTAAATCATTTAGAGATTAATTAATGAGAA 627
Db 207 IleIleGlnSerAlaAspAsnThrIYrValThrIYrLeuThrAspThrIYrShenThrIYs 226
Qy 628 AAACCTTTCTAAACCTTTCTTTCGACCGACGAGATTACCTGAGGAGAAATAGATTAT 687
Db 227 ValLeuAsnLys-----AspThrIYrIleAspGln 236
Qy 688 TTGGCTGACCAACAGAACGCCCTAAATGAGAGTGAAGTATTAAAGAACTTAAATAAC 747
Db 237 ValThrGlnProPheLeuValAlaLys-----Asn 246
Qy 748 AAATGTAGATGATTTCTTAATATGAGAGGCGCTGCAAAACAATATAGTCAATT 807
Db 247 GlnGlnGlnGlnGlnIYrPheSerGlnYAsnGlnIleMetLysYrYrAspGlnSerYr 265
Qy 808 ATAAAGAGATCAAAAAGGTTACGCT-----CCACAGACAGACGCGAAAGTATT 858
Db 266 Ile-----AsnAlaIYrIleIleProAspLeuYrThrGlnLeuIle 279
Qy 859 GGTACAGCCAGTGCAGTACCTGCAAAATTCGCAATTAATTAATTAATGAGTGAAGCTA 918
Db 280 -----SerLeuAlaYrIleuAspSerAsnLeuYrSerThrAsnLeu 293
Qy 919 AGCCAAAAGACAGGCTA---ACCTTACTCAAAATGATACGAAATACATGACGCTT 975
Db 294 ThrThrThrAsnIYrThrYrSerPheValAlaGluAsnThrSerAsnLysLeu---Val 312
Qy 976 CATAGTGT 984
Db 313 SerSerVal 315

RESULT 8
Q81SF7 PRELIMINARY; PRT; 908 AA.
AC Q81SF7;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DB Hypothetical protein.
GN ORFNames=PF1255C;
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
CX NCBI_taxid=36329;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22255705; Pubmed=12368864;
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA Carlton J.M., Pain A., Nelson K.B., Bowman S., Paulsen I.T., James K.,
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA Chan M.S., Nene V., Shallow S.J., Sun B., Peterson J., Angiuoli S.,
RA Pertea M., Allen J., Selengut J., Hult D., Mather M.W., Valdivia A.B.,
RA Martin D.M., Fairhead A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,

```

```

RA Venter J.C., Carucci D.J., Hofman S.L., Newbold C., Davis R.W.,
RA Fraser C.M., Barrell B.,
RA "Genome sequence of the human malaria parasite Plasmodium
RA falciparum".
RL Nature 419:498-511 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RA Hyman R.W., Fung E., Conway A., Kurd O., Mao J., Miranda M.,
RA Nakao B., Rowley D., Tamaki T., Wang F., Davis R.W.,
RL Submitted (Jan-2003) to the EMBL/Genbank/DBJ databases.
DR EMBL; AB014848; AAN6337.1; -.
KW Hypothetical protein.
SQ SEQUENCE 908 AA; 108485 MW; 48FEE8872F7C670 CRC64;

Alignment Scores:
Pred. No.: 2.67 Length: 908
Score: 113.00 Matches: 74
Percent Similarity: 36.66% Conservative: 51
Best Local Similarity: 21.70% Mismatches: 112
Query Match: 6.51% Indels: 104
DB: Gaps: 18

US-10-617-962-2 (1-1008) x Q81SF7 (1-908)
Qy 127 ATGAATATCGAGCTAAGAAATATTAATTAAGCTTACGCTTGGCTGATGATTCAT 186
Db 168 LeuYrIYrIleProLeuYrIleIleYrShen-----SerThrLeuHe 183
Qy 187 AATGTATCTAAACCTTCTGATGACTTATTAAGATAAAGAGACTGTGAGAAATTTAT 246
Db 184 AspIleSerAsnLeuYrAsnGlnIleYrYr----- 194
Qy 247 CAAGATATATGCTAATCTTTCATCTGACATTAATAGTGAATATGTCATCA----- 300
Db 195 -----TyrIleAsnGlnValIleLeuMetSerGlnYrAspGlnAsnIle 209
Qy 301 ATTCTAAATATGATGCAAAATGATTTTATTAAGATGCAATGATTTTGAAGTCAATAT 360
Db 210 LeuSerIYrValGlnIleIleYrAsnIleYrAsnAsnIleAsnSerGlnYrSerYr 229
Qy 361 CCTCAAAACATGGAATGTTCTGAGCTTGAATAAATCAATGATGATGCTTATTCAGAT 420
Db 230 AspHisLeuValThrSerYrLeuPheIleAspThrGlnAsn---AspIleYrSerAsp 248
Qy 421 GACGATTAATTAATGCA-----CTATATTTTCTCTGACAGAAATCCATGAG 474
Db 249 AspHisIleIleValGlnAspAsnAsnCySerPheThrThrThrGluYrIleAsnIleGln 268
Qy 475 GAATATCAACATCAAAATGCCGCAAGATTTTAAATTA----- 513
Db 269 AsnAsnLeuYr---AsnIleTrpAsnLeuLeuSerValHisIleHisAsnAsnIleCys 287
Qy 514 ---ATGATTTCTTATTAACCTTATCTGCTGTAACCTGCACTGGAGAGAGATTTTCA 570
Db 288 AsnValAspIleYrPheThr-----GlyLys-----LeuSer 298
Qy 571 AAAAATCTTTTCAATGATGATGAGGCTTAATCATTA-----GAGAAAT 612
Db 299 ProArgPheYrGlnIYrValAsnMetLysGlnLeuYrAspIleProAsnIleGlnYr 318
Qy 613 TATTTTGAAGAAAAAACTTTCAACCTTCTTCTTTCGACACGCGACAGATTAACCTGAT 672
Db 319 Tyr---GluAsnIleAsnPheSerSerLeuPheYrAsnPro---LysIleAsnAsp 336
Qy 673 GCGAATATGCTTATTTGGCTGACCAACAGAGCGCTTAATGAGAGTGAAGT----- 726
Db 337 GlnYrIleYrAsnIleValAsnGlnThr---IleProHisTrpSerValCysGlnGln 355
Qy 727 -----TTTAAACAATTTAAATTAACAATTCAGAAAT----- 759
Db 356 LysIYrIleGlnIleAspYrGlnYrGlnYrIleYrMetAspYrAsnAsnAsp 375
Qy 760 -----GATTTTCTAAT 771

```



```
Db 376 IleTyrTyrLysTyrAsnTyrPheLeuAspProTyrAsnLysLeuLysValSerAsn 395
Qy 772 ATGAAGGGGGCTGCAAAACAAAGTATAGTTCATTATTAAGAGGTACAAAGGGTAAAC 831
Db 396 ---AspAsnValHisArgAsnIleTyrSerSerLeuIleLysLys----- 409
Qy 832 GCTCCACAGACAGCAGCCGAAAGATTTGGTACAGCCAGTGGCAGTAACTGGAAAAATTG 891
Db 410 -----GluThrIle 412
Qy 892 CCGAATAATTTATATAGTGTGAGG-----CTAAGCCAAAGACAGGGTAAACC 939
Db 413 AsnAsnAsnLeuPheSerLeuAsnAspGluGlnLeuLeuLysGlnThrPheGlu 432
Qy 940 TTTACTCAAAATGATATACATATCAATATGACGGTTCATAGTGTGGAATCTATTAATA 999
Db 433 LysAsnLysAsnAspArgGlnAsnAsnAsnTyrAspAsnAsnAsnValIleHisTyrLys 452
Qy 1000 AAT 1002
Db 453 Asn 453

RESULT 9
Q9UN04 PRELIMINARY; PRT; 1822 AA.
ID Q9UN04
AC Q9UN04;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE ABC transporter, putative.
GN Name=PR0590w; Synonyms=MA1P3_03;
OS Plasmodium falciparum (Isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=2255708; PubMed=1236867;
RA Hall N., Pain A., Berriman M., Churcher C., Harris B., Harris D.,
RA Mungall K., Bowman S., Atkin R., Baker S., Barron A., Brooks K.,
RA Buckee C.O., Burrows C., Cherevach I., Chillingworth C.,
RA Chillingworth T., Christodoulou Z., Clark L., Clark R., Corton C.,
RA Crozin A., Davies R., Davis P., Dear P., Dearden F., Doggett J.,
RA Felwell T., Goble A., Goodhead I., Gwilliam R., Hamlin N., Hance Z.,
RA Harper D., Hauser H., Hornsby T., Holroyd S., Horrocks P.,
RA Humphrey S., Jagels K., James K.D., Johnson D., Kerhornou A.,
RA Knights A., Konfortov B., Kyes S., Larke N., Lawson D., Leonard N.,
RA Line A., Maddison M., McLean J., Mooney P., Moule S., Murphy L.,
RA Oliver K., Ormond D., Price C., Quail M.A., Rabinowitsch E.,
RA Rajandream M.A., Rutter S., Rutherford K.M., Sanders M., Simmonds M.,
RA Seeger K., Sharp S., Smith R., Squares R., Squares S., Stevens K.,
RA Taylor K., Tivey A., Unwin L., Whitehead S., Woodward J.,
RA Sultston J.E., Craig A., Newbold C., Barrell B.G.;
RT "Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13.";
RL Nature 419:527-531 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RA Murphy L., Pain A., Berriman M., Bowman S., Churcher C., Harris B.,
RA Harris D., Lawson D., Quail M., Rajandream M., Hall N., Barrell B.;
RL Submitted (Sep-1998) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: Belongs to the ABC transporter family.
EMBL, AL011746; CAB63558.1; -.
DR GO; GO:0016021; C:Integral to membrane; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. .; IEA.
DR GO; GO:0000166; F:nucleotide binding; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR011527; ABC_membrane_1.
DR InterPro; IPR001140; ABC_TM_transp.
DR InterPro; IPR003439; ABC_transporter.
DR Pfam; PF00664; ABC_membrane_1.
DR Pfam; PF00005; ABC_tran; 1.
```

```
DR SMART, SM00382; AAA; 2.
DR PROSITE; PS00929; ABC_TM1F; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER_1; UNKNOWN_1.
DR PROSITE; PS50893; ABC_TRANSPORTER_2; 2.
KW ATP-binding.
SQ SEQUENCE 1822 AA; 214439 MW; 04274BD126938F03 CRC64;

Alignment Scores:
Pred. No.: 3.03 Length: 1822
Score: 112.50 Matches: 77
Percent Similarity: 32.99% Conservative: 53
Best Local Similarity: 19.54% Mismatches: 127
Query Match: 6.48% Indels: 137
DB: 2 Gaps: 17

US-10-617-962-2 (1-1008) x Q9UN04 (1-1822)
Qy 37 TATCCACCCGTTGAAACCAAAATAGCAGACGATATAGTACGATCTA-----AACTTT 90
Db 759 TyrThrIleLeuGlnSerGlnLeuLeuAsnAspLeuSerThrIleGlnHisGlyAspMet 778
Qy 91 AAGCAAAAGATGAGGGTCATACAGCATCATATGGAATTTGATATGAGCTAAGAAATA 150
Db 779 LysTyrIleAsnAspAspHisAsnLeuSerLysGlnLysValArg-----Ile 795
Qy 151 ATATTAGCTTACGCTTTGGCTGTGAAGTGATTTTCATTAATGATCTTAACCTTCGATGAC 210
Db 796 CysLeuAlaArgAlaLeuTyrGlnHisTyrIleHis---MetHisLysLeuCysThrAsp 814
Qy 211 TATTATTAAGATTAAGAAGACGCTGAGAGGAATTTATCAAGAAATATATGCTTAATCTTCA 270
Db 815 TyrGlnLysLys----- 818
Qy 271 TCTGCACATTATGAGTAAATATGTCATCAATTTCTTAAGATATGCAATGCTTTTAT 330
Db 819 -----LeuIleGlnProAsnGlnIleLeuAspLysAspLeuIleAsn----- 832
Qy 331 AAGAAATGACTGGAATTTGAAAGTCAATATCTCTCAAAACATTTGGAATGTCCTGAGCTT 390
Db 832 ----- 832
Qy 391 GAAATTAACCATGAGTACGCTTATTCACATGACGATTAATTTATAGCACTATTTTTC 450
Db 833 ---AsnLysAsnIleSerSerTyrAsnAsnLysSerLysLeuValAsnTyr----- 849
Qy 451 TCTGTACAGCAAAATTCACCTGAGGAAATCAACATCAAAATGCCGCAAGA----- 501
Db 850 -----AsnIleProPheAsnGlnLysnTyrLeuGlnLysCysLeuMetAspAsn 866
Qy 502 -----TTTTTAATTAATGATTTCTTATTTACCTTATCTGCTGTAACCTTCACTGGGA 555
Db 867 AsnPheTyrLeuTyrLeuLeuAspAspIlePheThr---SerLeuAspProSerIleSer 885
Qy 556 AGAGGAGATTTTTCAAA----- 573
Db 886 LysLysIleLeuPheSerAsnLeuPheCysLysGlnAspAsnIleSerPheLysAspAsnCys 905
Qy 574 AACTTTCAATATGATTAAGAGGCTAATATCATTTAGCAATTAATTAAGAGAAAAACCTT 633
Db 906 SerPheIleIleSerMetAsnLysSerThrLeuAspAsnPheLeuIleGlnAspIleLeu 925
Qy 634 TCTAAACCTTTCTTTCGACACCGACAGATTAACCTGATGCGACAGATAGCTTATTTGGCT 693
Db 926 AspAsnValGlnTyr-----GluValAsnIlePheGlu 936
Qy 694 GGACCAACAGAAAGCCCTAAATGAGAGTGAATTTTAAGAAGCT-----AAAATTAAC 747
Db 937 IleGlnAspLysThrLeuLysTyrArgLysnIleSerGlnTyrMetGlnLysAsnAsn 956
Qy 748 -----AAATCTAGAAATGATATTTCTAATATGAA----- 777
Db 957 LeuAsnIleThrLysGlnSerHisTyrGlnLysSerAsnLeuAsnThrIleAspTyrThr 976
```


Qy	778	-----GGGGCGTCGACCAAAAGTATGTCATT	807
Db	977	ArglleYsLeuPhaSpGIuValGlueaEnHISVallyPhisSerAnlySwetlle	996
Qy	808	ATTAAGAAGTACAA-----AAGGTATACGCTCCACAGACAGCAGCAAAATATGCT	861
Db	997	TyrlYsGluValatYrPheValYlsGlyEnHthr-----GluSerValSerPheGluLeaP	1011
Qy	862	ACAGCCAGTGCAGTAACTCGAATAATG-----	891
Db	1016	SerlleaEnLySglutYrIleYsIlySwetLySlyLyAsnTyrlYsIySgluHISmet	1033
Qy	892	-----CCGAATAATTTATATGTCGAGGCTAACCCAAAAAGACAGCGTA	936
Db	1036	AsnLyAsaEnLyAsaPsaHnSAsnAsaEnSerAsnLyAsaPsaPHisIle	1055
Qy	937	ACCTTACTCAAAATGAT-----	954
Db	1056	AsnIleAsaEnLyAsaPsaHnSAsnAsaEnSerAsnLyAsaPsaPHisIle	1077
Qy	955	ACTGACAAATACAAAGACGGTTCATATGTTGAACATCATAT	996
Db	1076	ThraSapSerProThrValSerSerIleuGlyAsnGluTyrl	1089
RESULT 10			
ID	Q7RT20	PRELIMINARY;	PRT;
AC	Q7RT20;		995 AA.
DT	01-MAR-2004	(TREMBLrel. 26, Created)	
DT	01-MAR-2004	(TREMBLrel. 26, Last sequence update)	
DT	01-MAR-2004	(TREMBLrel. 26, Last annotation update)	
DE	Hypothetical protein.		
GN	Name=PY00181;		
OS	Plasmodium yoelii yoelii.		
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.		
OX	NCBI_TaxID=73239;		
LN	(1)		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=17XNL;		
RX	PubMed=12368865;		
RA	Carlton J.M., Angluoli S.V., Suh B.B., Kooij T.W., Perrea M.,		
RA	Silva J.C., Ermolaeva M.D., Allen U.E., Selengut J.D., Koo H.L.,		
RA	Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,		
RA	Shallow S.J., van Aken S.E., Redmiller S.B., Feldblyum T.V.,		
RA	Florens L., Quackenbush J.J., Sedegah M., Shoaihi A., Cummings L.M.,		
RA	Florens L., Yates F.R. III, Raine J.D., Sinden R.E., Harris M.A.,		
RA	Cunningham D.A., Preiser P.R., Bergman L.W., Valdiva A.B.,		
RA	van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,		
RA	Salberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,		
RA	Carucci D.J.;		
RT	"Genome sequence and comparative analysis of the model rodent malaria		
RL	parasite Plasmodium yoelii yoelii.";		
CC	Nature 419:512-519(2002).		
CC	-!- CAUTION: The sequence shown here is derived from an		
CC	EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is		
CC	preliminary data.		
DR	EMBL; AABL01000048; EAA21179.1; -.		
DR	InterPro: IPR001862; MAC_perforin.		
DR	Pfam: PF01823; MACPF; 1.		
KW	Hypothetical protein.		
SO	SEQUENCE 995 AA; 114106 MW; CC0889FD84EB63EA CRC64;		
Alignment Scores:			
	Pred. No.:	3.85	Length: 995
	Score:	111.00	Matches: 68
	Percent Similarity:	37.17%	Conservative: 45
	Best Local Similarity:	22.37%	Mismatches: 97
	Query Match:	6.40%	Indels: 94
	DB:	2	Gaps: 15
Qy	US-10-617-962-2 (1-1008) x Q7RT20 (1-995)		
	187 AATGATCTTAACCTTCGTATGATCTATAT-----AAGCAT 2222		

Db	615	AsnMetCysThrIlePheAspPheThrIleHisIleSylMetGlyLeuSerValIysAspArg	634
Qy	223	AAAGAGACTGCTGACGAGATTTATATCAAGAAATATATGTCTAATCTTCACTGACATTA	282
Db	635	LysGlnValGlnIlePheIleAsnIleHisIleSylSerIleThrValSerIleLeuGly	654
Qy	283	GGTGAAGAAATGGGATGCAAAATTTCTAAAGATATGGCAAAATGGTTTTTAATGAATGA	342
Db	655	GlnHisProGlyLeuAsnIleAspAsp	671
Qy	343	GATTTTGGAAGGTCATATATCTCAAAACAATTTGGAATGTCTCTGAGCTGGAATAATA	399
Db	672	AspSerIleAspLys	685
Qy	400	CCATTGAGCTGTATTATTCAGATGACGATAAATATTATTAAGCATATATTTTCTCTGA	456
Db	686	LeuProPheSerPhePheMetAspAspProAsnMetIleLysAlaIle	701
Qy	457	CAGGAATATCCATCGAGAGAAAATCAACAATCAATGCCGCAAGATTTTAAATTAATT	516
Db	702	AsnAspAlaLeuMetCpHeThrGlyLeu	710
Qy	517	GATTTCTATTATACCTTATCTGCTGTAACCTTACAGCGGAGAGAGATTTTTCACAAAAC	576
Db	711	ThrTyrGlyValGlnIlePheAspGlnAsn	720
Qy	577	TTTTTCAATGGAATTAGAGGCTTAATCATTTAGAGAAATATATTATTTAGAGAAAACCTTCT	636
Db	721	GlnTyrAsnAsnAsnGlnIleSerIleGlyAsnTyrLeuGlnLysSerIleGln	738
Qy	637	AAACCTTCTCTTGCACCAACCGCAGATTTACCTGATGCGAATAGT	690
Db	739	LysIleTyrHisIleGlySerProGlyLeuThrCysProIleGlySerThrIleLeu	758
Qy	691	GCTGAGCAACAAGAACGCGCTAAATGAGAGATGAGTTT	729
Db	759	MetGlyPheSerLeuAsnLeuAspPheTyrIleAsnGlnAsnLeuAsn	774
Qy	730	AAAGAACTTAATAATAATCAAAATCTAGGAATGATTT	765
Db	775	GlnIleIleGlyIleAsnIleAspGlnIleMetLysGlnSerCysSerGlyAsnGlyPhe	794
Qy	766	TCTAATATGAGAGGCGCTGCAAAACAAGATAGTTCATTT	807
Db	795	ThrAsnIleTyrSerAspIleArgIleTyrGlyLeuCysSerGlnIleTyrPhe	814
Qy	808	ATTAAGAGACGTACAAAGGCTAACGCTCCACAGACAGACGCG	849
Db	815	IleLysGlnValValGlnGlnAsnGlnSerThrLysThrIleAlaThrCysProGlnAsp	834
Qy	850	AAAAGATTGTGTACACCCATGCGACAGTAAC	879
Db	835	SerValIleLeuPheGlyPheAlaLeuMetLysGlyIleGlyArgSerSerIleAsnThr	855
Qy	880	CTGAGAAATATTCGCAATATTATTAATAGCTGAGGCTAAGCCAAAAGACAGGGTAACC	939
Db	855	ValAspLeuTyrProCysArgThrGlyGlnAsnSerCysSerAla	863
Qy	940	TTTACTCAAAAT	951
Db	870	ValLeuGlnAsn	873
RESULT 11			
081H6			
PRELIMINARY; PRT; 1794 AA.			
081H6			
01-MAR-2003 (TRMBLrel. 23, Created)			
DT 01-MAR-2003 (TRMBLrel. 23, last sequence update)			
01-MAR-2004 (TRMBLrel. 26, last annotation update)			
DE Hypothetical protein.			
GN ORFNames=PF11_0420;			
Plasmodium falcipterum (isolate 3D7).			


```

Query Match: 6.34% Indels: 102
DB: 2 Gaps: 13
US-10-617-962-2 (1-1008) x Q8II04 (1-3468)

QY 67 GATATAGTACCTAATCTTAATTAAGCAACAGATGAGGCTCATACGACATCATATGCA 126
   |||||
DB 872 AsnValIleHisAsnLysAsnPhelLysIleAsnGluIleAspHisPhelLysGluLysAsn 891

QY 127 --ATTGAATATCGAGTAAAGAAATATATTAGCTTACGCTTGGCTGAAGTGT--- 180
   |||
DB 892 ThrLeuSerLysGluSerLysAsnIleIleAsnThrCysLysAspAsnLeuGluLysAsp 911

QY 181 -----ATTCAATATGATCTAAACCTTCCTGATGAC----- 210
   |||||
DB 912 GluAsnAsnIleAsnAsnIleAsnGluLysAsnAsnSerLeuGlnPheAspLysAsn 931

QY 211 ---TATTATTAAGATTAAGAGACTGCTGAGAGAAATTATCAAGAAATAT--- 255
   |||||
DB 932 IleLeuLysAsnThrMetSerLeuGlnLysAspLysAsnIleLysAsnArgLys 951

QY 256 -----ATGCTAATCTT 267
   |||||
DB 952 AlaAsnThrSerAsnIleAspAsnIleAsnLysThrSerAsnAsnIleSerAsnLys 971

QY 268 TCATCTGCACTATTAGTGAATAATGTCATCAAAATTCT-----AAAGATATGACA 318
   |||
DB 972 SerAspIlePheIleAspAsnAsnLysSerLysSerLysThrAsnLysThrAspIle 991

QY 319 AATGCTTTTAAAGATGAACCTGATTTGAAGTCAATATCTCAAAACATTTGGAAAT 378
   |||||
DB 992 AsnIlePheSerLysAsnLysSerAsnAspGluAspThrPheSerLysCysPheThrTyr 1011

QY 379 GTTCTGAGCTTGAATAAACCATGAGTCTTATTTAGATGAGATTAATTTATACCA 438
   |||||
DB 1012 LysGluHisLeuSerAsn-----TyrAsnLysAsnAspGluAsnLeu 1026

QY 439 CTATATTCTTCTGTGACAGAAATTCACCTGAGCAAGAAATCAACATCAATCCGCA 498
   |||
DB 1027 ThrThrPheSerAsnAlaPheHisIleProLysIleAsnAsnIleLysSer----- 1044

QY 499 AGATTTTAAATTAATGATTTCTTATTTACCTTATGCTGTAACCTCACTGGAAG 558
   |||
DB 1045 -----ThrHisAsnAspIleLeuAsnIleSerAsn 1054

QY 559 AGCATTTTCAAAAACTTTTACATGATTAAGGCTTAATCTTTGAGATTAATTT 618
   |||||
DB 1055 LysTyrTyrAsnLysAsn-----AsnThrAsnLysThrAsnIleAspIlePheGln 1072

QY 619 GAGAGAAAAAACTTTCTAAACCTTTCTTGACACCGCAGAGATTACCTGATGCGACA 678
   |||
DB 1073 AsnLysAsnSerLeuAspMetPro----- 1080

QY 679 ATAGGTTATTGGCTGACCAAGAACGCTTAATGAGAGTGATTT----- 729
   |||
DB 1081 -----GluThrAsnLeuIleLysGluLysGlnPheIleLysThr 1094

QY 730 -----AAAGAACTTAATAATCAAAATCTAGAGATGATTTCTAAT 771
   |||||
DB 1095 AspMetSerSerLysAspAsnSerLysAsnAspAsnSerArgAspIleArgGluAsn 1114

QY 772 ATGGAAGGGGCTGCAAAACAAAGATAGTTCATTTATTAAGAGAGTCAAAAGGTTAC 831
   |||||
DB 1115 IleAspCysSerValLysAsnGluLysGlnSerPheAsnLysAsnMetSerAsnLysAsn 1134

QY 832 GCTCCACAG----- 840
   |||||
DB 1135 LeuGlySerMetAsnValSerProCysAsnHisMetLysThrAsnGluAsnAsnAspIleIle 1154

QY 841 ACAGACGGAAGAAAGTATTGTTGACAGCCAGTGGCAGTAACTTGGAATA-----TTGCCG 894
   |||||
DB 1155 AsnIleSerLysTyrIleGluHisArgIleArgLysAsnAlaGluGluAsnThrMetLys 1174

```

```

QY 895 AATAATTATTAATAGTGTGAGGCTTAAGCCAAACAGCGGTAACCTTTACTCAAAATGAT 954
   |||||
DB 1175 AsnAsnIle---AsnGluAspThrSerLysAspPheAsnCysPheSerAsnAsnGlu 1193

QY 955 ACTGACAATCAATGACGGTTCATAGTGT 984
   |||||
DB 1194 LysLysLysAsnPheThrThrAsnAsnIle 1203

RESULT 13
Y028 BORBU STANDARD; PRT; 414 AA.
ID Y028 BORBU
AC O50959;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-OCT-2004 (Rel. 45, Last annotation update)
DE Hypothetical ANK-repeat protein BB28.
GN OrderedLocustNames=BB28;
OS Borrelia burgdorferi (Lyme disease spirochete).
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia.
OX NCB1_TaxID=139;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35210 / B31;
RX MEDLINE=98065943; PubMed=9403685; DOI=10.1038/37551;
RA Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,
RA Ladhira R., White O., Ketchum K.A., Dodson R.J., Hickey E.K.,
RA Gwinn M.L., Dougherty B.A., Tomb J.-F., Fleischmann R.D.,
RA Richardson D.L., Peterson J.D., Keilavage A.R., Quackenbush J.,
RA Salzberg S.L., Hanson M., Van Vugt R., Palmer N., Adams M.D.,
RA Gocayne J.D., Weidman J.F., Utterback T.R., Matthews L., McDonald L.A.,
RA Artchak P., Bowman C., Garland S.A., Fujii C., Cotton M.D., Horst K.,
RA Roberts K.M., Hatch B., Smith H.O., Venter J.C.;
RT "Genomic sequence of a Lyme disease spirochete, Borrelia
RT burgdorferi."
RL Nature 390:580-586 (1997).
CC -!- SIMILARITY: Contains 2 ANK repeats.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: AE000792; AAC66335.1; -.
CC PIR: H70219; H70219.
CC TIGR: BB28; -.
CC InterPro: IPR002110; ANK.
CC Pfam: PF00023; ANK, 2.
CC SMART: SM00248; ANK, 2.
CC PROSITE: PS50084; ANK_REPEAT, 1.
CC PROSITE: PS50297; ANK_REPEAT_REGION, 1.
CC ANK repeat; Complete proteome; Hypothetical protein; Plaemid; Repeat.
FT REPEAT 326 355 ANK 1.
FT REPEAT 359 389 ANK 2.
SQ SEQUENCE 414 AA; 49379 MW; BEACASD8PA26680 CRC64;

Alignment Scores:
Pred. No.: 4.82 Length: 414
Score: 109.50 Matches: 72
Percent Similarity: 36.02% Conservative: 44
Best Local Similarity: 22.36% Mismatches: 98
Query Match: 6.31% Indels: 108
DB: 1 Gaps: 14

US-10-617-962-2 (1-1008) x Y028 BORBU (1-414)

QY 10 CAATTAACCTGATGATAGAAAGTGATATCCACCCGTTGAAAGCAATATGACAGAGAT 69
   |||||
DB 178 GluIleSerIleAspGluAsnGlyAsnIleIleProValIleLysAsnThrAsnLysGln 197

```

```
QY 70 ATAGTACGTATTA-----AACTTAAGCAACAGATGAGGTATACAGA 117
Db 198 ILeThrgluLeuLeuPheThyAspGlnuThyAsnLeuLeuProPheLysPheGlu 217
QY 118 TCATATGAAATGATAT-----CGAGCTAAGAAATA 150
Db 218 SerTyrLysValArgPheAspLeuLeuProGluAsnLysAsnLeuAsnPheLysGlu 237
QY 151 ATATTAGCTTACGCTGCTGTAAGTGTATTCATATGATCTAACTTCGATGAC 210
Db 238 ILeAsnValTyrTyrLeuAspGlnuLysAsnLeuLeuThr-----ProLeuGlu 253
QY 211 TATTATGAAATTAAGAGAGCTGAGAGATTTTCAAGATATATGCTATATCTTCA 270
Db 254 TyrTyrLysAsnAsnLeuAspMetSerProTyrTyrLeuAsp----- 267
QY 271 TCTGACATATTAGGAAATATGATCAAAATTTCTAAAGATATGCAAAATGTTTAT 330
Db 268 -----LeuGlnGlnuAsnLysAspAsp-PheLeuLys----- 277
QY 331 AAGAAATGAATCTGATTTGAAAGTCAATATCTCAAAACATTGGAAATGCTGAGCTT 390
Db 278 -----AlaIleLysIleLysLysGlnuTyrGlnuTyrIleGlnuLys 291
QY 391 GAAATTAACCATTTGAGTCTTATTCAGATGACGATTAATTTTACACTATATTTTTC 450
Db 291 sLys--LysGlnLeuGlnuAsnLeuThrgluAsnAspLysLeu-----Asp 305
QY 451 TCTGTACAGAAATTCACCTGAGGAAATCAACATCAAAATGCCGAATTTTAA 510
Db 306 AspPheLysGlnuPheLeuLeuLysAsnAsn----- 316
QY 511 TTAATTTGATTTCTTATTTACCTTATCTGCTGCTTACTGCACTGGAAGATTTTCA 570
Db 317 -----IlePheSerLeuAsnThrxIlePheSerAsnGlyAsnProIlePheThx 332
QY 571 AAAAATTTCATGATGATTTAGAGCTTAATCATTTAGAGATTTATTTAGAGAAAA 630
Db 333 -----TyrAlaIleAsnValLysAlaLysSerIleIleAsnTyrLeuLeuThrxLysGln 350
QY 631 CTTTCTAATACCTTTCTTTCGACACCGAGATTTACTGAGGAGATATTTTG 690
Db 351 PheAsn----- 352
QY 691 GCTGACCAACAGAACGCTTAATGAGAGTGTATTTAAAGAACTTAATAACAA 750
Db 353 -----IleAsnLeuThx 356
QY 751 TCTAGGAATGATTTTCTATATTTGAGAGGCTGCAAAACAAAGTATGCT--TCATTT 807
Db 357 AsnGlnAsnSerGlnThrxIleAsnLysSerAlaIleIleGlnLysTyrAspLeuAsnPhe 376
QY 808 ATAAAGAG--GTACAAAGGTTAACGCTCCACAGACGACGCAAGAAAGTATTTGTACA 864
Db 377 ILeLysSerLeuIleGlnuLysGlyAlaAsnPro-----AsnIle 389
QY 865 GCCAGTGCAGTACCTGGAATAATTTGCCGAATTAATTTATAGTGTGAGCTTAAGCCA 924
Db 390 ArgAspGlyAspAsn-----LysLeuProIleAsp--TyrSerAspLysThxSerGln 406

RESULT 14
Q7RN55 PRELIMINARY; PRT; 1491 AA.
AC Q7RN55;
DT 01-MAR-2004 (TREMBLrel. 26, Created)
DT 01-MAR-2004 (TREMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
DE Hypothetical protein.
GN Name=PY01969;
OS Plasmodium yoelii yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
NCBI_taxonomy=73239;
RN [1]
```

```
RP SEQUENCE FROM N.A.
RC STRAIN=17XNL;
RX PubMed=12368865;
RA Carlton J.M., Angiolini S.V., Suh B.B., Koof T.W., Pertea M.,
RA Silva J.C., Ermolaeva M.D., Allen J.B., Selengut J.D., Koo H.L.,
RA Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,
RA Shalom S.J., van Aken S.B., Riedmuller S.B., Feldblyum T.V.,
RA Cho J.R., Quackenbush J., Sedegah M., Shoabdi A., Cummings L.M.,
RA Florens L., Yates F.R. III, Raine J.D., Sinden R.E., Harris M.A.,
RA Cunningham D.A., Preiser P.R., Bergman L.W., Valdivia A.B.,
RA van Lin J.H., Jansz C.J., Waters A.P., Smith H.O., White O.R.,
RA Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,
RA Carninci D.J.,
RT "Genome sequence and comparative analysis of the model rodent malaria
RT parasite Plasmodium yoelii yoelii.",
RL Nature 419:512-519(2002).
CC -! CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL, AABL01000539; EAA21375.1; -.
KW Hypothetical protein.
SQ SEQUENCE 1491 AA; 175817 MW; 7FA33125E6B1F26E CRC64;
```

Alignment Scores:

Pred. No.:	Length:	Matches:
Score: 5.15	1491	76
Percent Similarity: 109.50	Conservative: 57	
Best Local Similarity: 41.96%	Mismatches: 104	
Query Match: 23.97%	Indels: 80	
DB: 6.31%	Gaps: 17	

US-10-617-962-2 (1-1008) x Q7RN55 (1-1491)

```
QY 187 AATGATCTAACTTCTGATGATCTATTAAG-----AATAAAG 228
Db 11 SerLeSerLysLysTyrGlnuAspAspTyrLysAlaGlnuLeuThrglnuAsnGlnu 30
QY 229 ACTGCTGAGAAATTTATCAAGAA-----TATATGCTATCTT--TCATCTGCA 276
Db 31 AsnAlaLys-----TyrLysGlnuAspAsnAlaTyrGlnuThrxAsnIlePheGlnuAsnSer 48
QY 277 CTATTAGGTGAA--AATGATGATCAATTTCTAAAGTATGCAAAATGCT--TTTAT 330
Db 49 ILeHisProGlnuIleLysAsnAsnLysPheProLeuAspTyrTyrAsnSerGlnuGlyThx 68
QY 331 AAGATGAATGATTTT-----GAGGCTAATATCTCAAAACATTTGGAATGCTCT 384
Db 69 LysAsnGlnuAlaAspTyrMetGlnuLysAsnLysTyrGlnuSerCysIleAspAsnLysGly 88
QY 385 GAGCTGAATAATTAACCA-----TTGAGTGCTTATTCAGATGAGATTAATTA 432
Db 89 AsnSerMetAsnLysProAsnMetGlnuThrxPheGlnuAsnIleAsnAspHisGlnuLys--- 107
QY 433 TTAGCATATATTTTCTCTGTACAGAAATTCACCTGAGGAGAAATCAACATCAAT 492
Db 108 -----IleGlnuAsnLysIleLysAsn 115
QY 493 GCCCGAATTTTAAATTAATGATTTCTTATTTACCTATCTGCTGTAATCTCACTG 552
Db 116 GlyAspArg-----SerIleSerGlnuSerIleAsnTyr 126
QY 553 GGAAGAGATTTTTCAAAAAATTTTACAATGATTAAGGCTTAATCTTAAG-- 609
Db 127 AsnLysGlnuAsnThrxGlnuMetAsnLeuLeuAsnIleLysGlnuAsnLysThrxGlnu 146
QY 610 -----ATTATATTTAGAGAAAAAATTT-----CTAAACTTTCTTTCGACACCG 657
Db 147 LysThrxAsnPheGlyAsp-AsnLysAsnPhAsnLeuLeuLysLysLysLeuThrxHisAs 166
QY 658 CAGAGATTAACCTGATGCGAGATAGTATTTGGCTGAGCAACAGAGCGGCTTAATG 717
Db 166 pGlnuAspAspLeuLeuAspAspLeuLeuLysTyrAsnAsnAsnGlnuLysGlnuLysAsnThx 186
```

Qy	718	AGAGCGAGTTTAAAGACTTTAAATAATACAAATCTAGAAATGATTTCTTAATATGGA	777
Db	186	rSerPheArgIlySlySlySlySlySlySgluIasnLeuAsgIuAanTyrIleSerSerIy	206
Qy	778	GGCGCTGCAAAACAAAGATAGTATGTTATTTATTAAGA-	-AG 81.6
Db	206	sIleasnSerGluLeuSerIleIleuAspIeuArgIlySlyAsnThrGluIleIyThrIy	226
Qy	817	GTACAAAAGGGTACAGCTCCACAGACAGCAGCGAAAGATTTGGTACAGCCAGTGGCAGT	876
Db	226	sPheGluIlySleuSerMetAsnSerGlnIlySlyS-	Gln-SerAspLeu 241.1
Qy	877	AACTGTGAAAAATTGCCGATTAATTTA-	-TAACTGTG 912.2
Db	242	AsnIleArgIyIleGlnAsnIlyThrSerPheIleAsnAsnThrThrAsnHisIleVal	261
Qy	913	AGCGTAAAGCCAAA-	-GACGGGTAACTTAACT 945.5
Db	262	AsnIleuLeuGlnIlySlyAsnSerGluGluIuIleMetIleGlnArgIleValIleGln	281
Qy	946	CAAAATGATCTAGACATCAATG----	-ACGGTCAATGTTT 984
Db	282	GlnSerAspThrAspAspIleValGlnIlySlyValHisThrIle	296

Result	15	07RLZ8	PRELIMINARY;	PRT;	1119 AA.
ID	07RLZ8				
AC	07RLZ8;				
DT	01-MAR-2004 (TrEMBLrel. 26, Created)				
DT	01-MAR-2004 (TrEMBLrel. 26, Last sequence update)				
DT	01-MAR-2004 (TrEMBLrel. 26, Last annotation update)				
DE	Hypothetical protein.				
GN	Name=PY02390;				
OS	Plasmodium yoelii yoelii.				
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.				
NCBI	Unlabeled				
NCBI	_TaxId=73239;				
LN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=17XNL;				
RX	PubMed=12368865;				
RA	Carlton J.M., Angiolini S.V., Suh B.B., Koolif T.W., Perea M.,				
RA	Saliva J.C., Ermolaeva M.D., Allen J.E., Seleguent J.D., Koo H.L.,				
RA	Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,				
RA	Shalom S.J., van Aken S.E., Riedmuller S.B., Feldblyum T.V.,				
RA	Chio J.K., Quackenbush J., Selegue M., Shoalib A., Cummings L.M.,				
RA	Florens L., Yates F.R. II, Raine J.D., Sindén R.E., Harris M.A.,				
RA	Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B.,				
RA	van Lin L.H., Jense C.J., Waters A.P., Smith H.O., White O.R.,				
RA	Salzberg S.L., Venter J.C., Frazer C.M., Hoffman S.L., Gardner M.J.,				
RA	Carucci D.J.;				
RT	"Genome sequence and comparative analysis of the model rodent malaria				
RT	parasite Plasmodium yoelii yoelii.";				
RL	Nature 419:512-519(2002).				
CC	-!- CAUTION: The sequence shown here is derived from an				
CC	EMBL/Genbank/DBD whole genome shotgun (WGS) entry which is				
CC	preliminary data.				
CC	EMBL; AAB0100656; EAA21832.1; -				
DR	InterPro; IPR011591; Botulinum.				
DR	Prodom; PD001963; Botulinum; 1.				
DR	Hypothetical protein.				
SO	SEQUENCE 1119 AA; 131470 MW; E3A1D182CC68A079 CRC64;				

Alignment Scores:	
Pred. No.:	5.55
Score:	109.00
Percent Similarity:	33.69%
Best Local Similarity:	20.16%
Query Match:	6.28%
DB:	2
Length:	1119
Matches:	76
Conservative:	51
Mismatches:	142
Indels:	108
Gaps:	17

US-10-617-962-2 (1-1008) x Q7RLZ8 (1-1119)

QY 7 ATACAATTACACCTGATGATAGAAAGTGGATATCCACCCGTTGAAAAGCAATAGCAGGA 66

Db	751		-----	-HistHraenLeuAlaGlu	765
Oy	67	GATATAGTACGATTAACAACTTAAAGCAAAACAGATGAGGCTCATACATCATATGGA	126		
Db	766	ThrlleValAsnlyslleSerAaplyGluAsnAsnGluGlyAsnAaplleLeuileAsn	785		
Oy	127	ATTGAATATGAGCTAAGAAAATAATATATAGCTTACCGCTTGGCTGTAAAGCTATTCAAT	186		
Db	786	TyrlLutrrp	788		
Oy	187	AATGTATCTAAACTTCCGTGATGACTATTTATAGATATAAGAGACTGCTGAGAAATTTAT	246		
Db	789	-----TylllyAsnlysglu-----	794		
Oy	247	CAAGAAATATATGCT--AATCTTCACTCGCACTATTAAGTGAATGCGTAATCAATT	303		
Db	795	---lysrTyrmethIstleSerleuLyAsnlyslleAapnsenSerAsnGluSerGluile	813		
Oy	304	TCT--AAAGATATGGCAAT--CGTTTATATAGAAATGAACTGGAT----	345		
Db	814	ThrglyAsnAapAlaLyAsnValleuileAlaIstHrleuAsnLybleuAapProkrg	833		
Oy	346	-----TTTGAAGTCAATATCTCAAAACATTTGGAATGTCTCGAGCTT	390		
Db	834	IlleAspAsnlysaenValProGluAenThrProSerAsnethPraen-----	849		
Oy	391	GAATAATAAACCATTAAGTGTCTTATTCAGATGACAGATAAATATTAAGCATATATTTTC	450		
Db	850	---AsnAsnAapleuAenTyrlleThraAsnlyslleSerleuAapSerleuHstlyrAsn	868		
Oy	451	TCGTGTACAGAAATTCACATGGAGAAAT-----CAACATCAAT	492		
Db	869	SerAsnTySerAapAsnIleSerAryAsnLeuAsnGlylleSerleuAsnAsnSerGlu	888		
Oy	493	GCCCAAGATTTTAAATTAATTAATGAT-----TTCTTA	525		
Db	889	AlaIstGluileCylySerSerleuAenGlnGluCyAsnAsnGlnAenGluThrPhehe	908		
Oy	526	TTTACCTTATGTGCTGTAACTTCACTGGAGAGAGATTTTTCAAAACCTTTACAT	585		
Db	909	GlnProIleuLyslleAapThrAapLeuGlySerAryGlysglyAlaLeuAsn-----	925		
Oy	586	GGATTAGAGCGTAATCAATTAAGATATATATAGAGAAAAAAAATTTCAACCTTTC	645		
Db	926	-----GluValGluAsnIleuAsnlysaenlySerProGluIleSerlle	940		
Oy	646	TTTGCACACCGCAGAGATTACTGTATGGCAGATAGGTTATTTGGCTGCACCAACA--	702		
Db	941	SerThrGluSerGlnlybleuValAapGlnThrlyA-----SerGlyGlnThrlys	957		
Oy	703	-----GAACCGCTTAATGAGAGATGAGCTTTTAAAGACTT	738		
Db	958	AsnlysaenThrIleleuCyAsnAapAlaVallylserTyrlleCyethThrAsnAsnAsn	977		
Oy	739	AAAAATTAACAATTAAG-----AATGGATTTTCTAATATAGAGAGGGGCTGCA	786		
Db	978	LyAsnAsnAsnAsnAsnAsnlysaenAsnSerAsnSerleuAenThrGlnlyleVal	997		
Oy	787	AAACAAAAGTAT-----AGTTCATTTATTAAGAAGTACAAAAGGGTAAACGCT	834		
Db	998	AsnGlnIleTyrlleAsnAsnAsnSerleuLeu-----AsnAla	1011		
Oy	835	CCACAGACAGAGGAAAAGTATTCATACGCCCATGGCACTAATCCTGAAAATTCGCC	894		
Db	1012	ProAsnLeuAlaAsnAsnIleLeuThrTySerSerThrGlyAsnTyrllyslThrIleSer	1031		
Oy	895	AATATATTTATATAGTGAAGGCTTAAGCCAAAAMAGACAGGCTAACCTTTACTCAAAATGAT	954		
Db	1032	AsnAsnThrIleAsnLeu-----AapSerIlelyshIleValAapGlnlle	1046		
Oy	955	ACTGACATACATAGACGGTTCATATAGTGTGGAATCTATTATAAATATATA	1005		

955 ACTGACATATCAATGACGGTTTCATAGTGTGGAACTCATTTATAAATATA 1005

Db 1047 LysGlnGluAsnIleLysIleHisMetGluProSerHisIleSerAsnIle 1063

Search completed: November 21, 2004, 08:54:00
Job time : 161.5 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 21, 2004, 07:15:36 ; Search time 579 Seconds
(without alignments)
9407.344 Million cell updates/sec

Title: US-10-617-962-2

Perfect score: 1008
Sequence: 1 atggttatacaattacacc.....ctcattataaataatgta 1008

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 3627888 seqs, 2701811610 residues

Total number of hits satisfying chosen parameters: 7255776

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications NA:*

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
5: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
6: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
18: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
19: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
20: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
21: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1008	100.0	1008	US-10-617-962-2	Sequence 2, Appli
2	663.8	65.9	1107	US-10-617-962-1	Sequence 1, Appli
3	663.8	1205	16	US-10-617-962-5	Sequence 5, Appli
4	50.6	5.0	188971	US-10-235-192A-27	Sequence 27, Appli
5	49.8	4.9	11394	US-10-240-453-95	Sequence 95, Appli
6	48.8	4.8	1866	US-10-349-680-73	Sequence 73, Appli
7	48.2	4.8	2767	US-10-301-533-23	Sequence 23, Appli
8	47.6	4.7	16217	US-10-311-455-997	Sequence 597, Appli
9	47.4	4.7	942	US-08-781-986A-531	Sequence 531, Appli
10	47.4	4.7	942	US-10-329-624-531	Sequence 531, Appli
11	47.4	4.7	158001	US-10-211-179-11	GENERAL INFORMATION
12	47	4.7	549	US-09-991-936-879	Sequence 879, Appli

13	47	4.7	11052	US-10-257-166-68	Sequence 68, Appli
14	46.4	4.6	260	US-10-424-599-100851	Sequence 100851, Appli
15	46.2	4.6	34722	US-10-322-881-700	Sequence 700, Appli
16	45.6	4.5	18283	US-10-321-613-326	Sequence 326, Appli
17	45.4	4.5	235070	US-10-087-192-1990	Sequence 1990, Appli
18	45.2	4.5	4985	US-10-094-240-10	Sequence 10, Appli
19	45.2	4.5	4985	US-10-056-405-10	Sequence 10, Appli
20	45.2	4.5	13377	US-10-311-455-1436	Sequence 1436, Appli
21	45.2	4.5	13377	US-10-221-714A-198	Sequence 198, Appli
22	45	4.5	6103	US-10-311-455-1664	Sequence 1664, Appli
23	44.8	4.4	1046	US-10-027-632-10421	Sequence 10421, Appli
24	44.8	4.4	1046	US-10-027-632-10421	Sequence 10421, Appli
25	44.8	4.4	3399	US-10-793-639-600	Sequence 600, Appli
26	44.6	4.4	973	US-10-424-599-58879	Sequence 58879, Appli
27	44.6	4.4	1488	US-09-815-242-4589	Sequence 4589, Appli
28	44.6	4.4	1491	US-09-815-242-8559	Sequence 8559, Appli
29	44.6	4.4	1491	US-10-282-122A-8174	Sequence 8174, Appli
30	44.6	4.4	13508	US-08-781-986A-120	Sequence 120, Appli
31	44.6	4.4	13508	US-10-329-624-120	Sequence 120, Appli
32	44.6	4.4	3673778	US-10-312-841-1	Sequence 1, Appli
33	44	4.4	609	US-10-424-599-102915	Sequence 102915, Appli
34	44	4.4	18683	US-10-311-455-285	Sequence 285, Appli
35	44	4.4	18683	US-10-240-452-33	Sequence 33, Appli
36	44	4.4	3673778	US-10-312-841-2	Sequence 2, Appli
37	43.8	4.3	54775	US-10-741-601-5659	Sequence 5659, Appli
38	43.8	4.3	3673778	US-10-312-841-1	Sequence 1, Appli
39	43.6	4.3	6418	US-10-311-455-296	Sequence 296, Appli
40	43.6	4.3	6418	US-10-221-613-30	Sequence 30, Appli
41	43.4	4.3	3885	US-10-377-636-3	Sequence 3, Appli
42	43.4	4.3	4985	US-10-094-240-10	Sequence 10, Appli
43	43.4	4.3	4985	US-10-056-405-10	Sequence 10, Appli
44	43.4	4.3	11805	US-10-311-455-1721	Sequence 1721, Appli
45	43.2	4.3	564	US-10-437-963-7175	Sequence 7175, Appli

ALIGNMENTS

RESULT 1
US-10-617-962-2
; Sequence 2, Application US/10617962
; Publication No. US20040055036A1
; GENERAL INFORMATION:
; APPLICANT: Commonwealth Scientific and Industrial Research Organisation
; APPLICANT: EAST, Peter David
; TITLE OF INVENTION: Toxin Genes from the Bacteria Xenorhabdus nematophilus and Photorhabdus luminescens
; FILE REFERENCE: 050179-0076
; CURRENT APPLICATION NUMBER: US/10/617,962
; CURRENT FILING DATE: 2003-07-14
; PRIOR APPLICATION NUMBER: US/09/463,048A
; PRIOR FILING DATE: 2002-12-13
; PRIOR APPLICATION NUMBER: PCT/AU98/00562
; PRIOR FILING DATE: 1998-07-17
; PRIOR APPLICATION NUMBER: PO 8088
; PRIOR FILING DATE: 1997-07-17
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 2
; LENGTH: 1008
; TYPE: DNA
; ORGANISM: Photorhabdus luminescens
US-10-617-962-2

Query Match 100.0%; Score 1008; DB 16; Length 1008;
Best Local Similarity 100.0%; Pred. No. 1.3e+198;
Matches 1008; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGTTATTAACAATTACACCGATGATGATGATTCACCGCTGAAGAACAATA 60
DB 1 ATGGTTATTAACAATTACACCGATGATGATGATGATTCACCGCTGAAGAACAATA 60
QY 61 GCAGAGATATATAGTATGATTAATACTTAAGCAACAGATGAGGTCTATACAGATCA 120


```
Db 61 GAGAGAGATAGTACGTAATCTTAACCTTAAGCAAGAGAGGTATACAGATCA 120
Qy 121 TATGGAATGATATCGAGCTAAGAAAATATATATAGCTTACGCTTGGCTGAAGTGT 180
Db 121 TATGGAATGATATCGAGCTAAGAAAATATATATAGCTTACGCTTGGCTGAAGTGT 180
Qy 181 ATTCAATATGATCTTAACCTTCCGTGATGACATATATATAGATTAAGAGACTGCTGAGA 240
Db 181 ATTCAATATGATCTTAACCTTCCGTGATGACATATATATAGATTAAGAGACTGCTGAGA 240
Qy 241 ATTATGCAAGATATATATGCTATATCTTATATCTGACATTTAGTGCAAAATGCTATCA 300
Db 241 ATTATGCAAGATATATATGCTATATCTTATATCTGACATTTAGTGCAAAATGCTATCA 300
Qy 301 ATTTTCAAAAGATATGCAAAATGCTTTTATTAAGAAATGCAATTTTGAAGCTCAATAT 360
Db 301 ATTTTCAAAAGATATGCAAAATGCTTTTATTAAGAAATGCAATTTTGAAGCTCAATAT 360
Qy 361 CCTCAAAACATTTGGAAATGCTTCCTGAGCTTGAATAATCAATGAGCTTATTCAGAT 420
Db 361 CCTCAAAACATTTGGAAATGCTTCCTGAGCTTGAATAATCAATGAGCTTATTCAGAT 420
Qy 421 GACGATTAATTTATGACATATATTTTCTGCTACAGAAATTCACATGAGGAAAT 480
Db 421 GACGATTAATTTATGACATATATTTTCTGCTACAGAAATTCACATGAGGAAAT 480
Qy 481 CAACATCAAAATGCGCAAGATTTTATTAATTAATGATTTCTATTTACCTTATCTGCT 540
Db 481 CAACATCAAAATGCGCGCAAGATTTTATTAATTAATGATTTCTATTTACCTTATCTGCT 540
Qy 541 GTAACTTCACTGCGAAGAGATTTTTCACAAAACCTTTTACAAATGATTAAGGCTTAA 600
Db 541 GTAACTTCACTGCGAAGAGATTTTTCACAAAACCTTTTACAAATGATTAAGGCTTAA 600
Qy 601 TCATTAAGAAATTTATTTAGAGAAAACCTTTCTTAAACCTTTCTTGGACCAACGCGAG 660
Db 601 TCATTAAGAAATTTATTTAGAGAAAACCTTTCTTAAACCTTTCTTGGACCAACGCGAG 660
Qy 661 AGATTACCTGATGCGAGATATGCTTATTTGGCTGACCAAGAAAGCGCTTAAATGAGA 720
Db 661 AGATTACCTGATGCGAGATATGCTTATTTGGCTGACCAAGAAAGCGCTTAAATGAGA 720
Qy 721 GTGAGTTTAAAGAACTTAAATAATCAATCTAGGAATGATTTTCTAATAAGAGG 780
Db 721 GTGAGTTTAAAGAACTTAAATAATCAATCTAGGAATGATTTTCTAATAAGAGG 780
Qy 781 GGTGCAAAACAAAGATATGCTTATTTAAAGAGGTAAGAGGTAAGCTCCACAG 840
Db 781 GGTGCAAAACAAAGATATGCTTATTTAAAGAGGTAAGAGGTAAGCTCCACAG 840
Qy 841 ACAGCAGCAAAAGATATGCTTACAGCAGTGGCAGTACCTGCAAAATTTGCGAATAT 900
Db 841 ACAGCAGCAAAAGATATGCTTACAGCAGTGGCAGTACCTGCAAAATTTGCGAATAT 900
Qy 901 TTATATAGTATGAGGCTTAAAGCAAGAGTAACCTTTACTCAAAATGATATCTGAC 960
Db 901 TTATATAGTATGAGGCTTAAAGCAAGAGTAACCTTTACTCAAAATGATATCTGAC 960
Qy 961 AATAAGATAGCGTTCTATGCTTGGAACTCATTTAATAAATATATGA 1008
Db 961 AATAAGATAGCGTTCTATGCTTGGAACTCATTTAATAAATATATGA 1008

RESULT 2
US-10-617-962-1
; Sequence 1, Application US/10617962
; Publication No. US20040055036A1
; GENERAL INFORMATION:
; APPLICANT: Commonwealth Scientific and Industrial Research Organisation
; APPLICANT: EAST, Peter David
; TITLE OF INVENTION: Toxin Genes from the Bacteria Xenorhabdus nematophilus and Photoc
; TITLE OF INVENTION: Luminescens
```

```
; FILE REFERENCE: 050179-0076
; CURRENT APPLICATION NUMBER: US/10/617,962
; CURRENT FILING DATE: 2003-07-14
; PRIOR APPLICATION NUMBER: US/09/463,048A
; PRIOR FILING DATE: 2002-12-13
; PRIOR APPLICATION NUMBER: PCT/AU98/00562
; PRIOR FILING DATE: 1998-07-17
; PRIOR APPLICATION NUMBER: PO 8088
; PRIOR FILING DATE: 1997-07-17
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1107
; TYPE: DNA
; ORGANISM: Xenorhabdus nematophilus
US-10-617-962-1

Query Match 65.9%; Score 663.8; DB 16; Length 1107;
Best Local Similarity 79.4%; Pred. No. 2.1e-127;
Matches 835; Conservative 0; Mismatches 162; Indels 54; Gaps 2;

Qy 12 ATTAACACTGATAGTAGAGATATCCACCCGTTGAAAAGCAATATGACAGATAT 71
Db 57 AGTACGCTGATGATATAGTGAATATCAACCCGTTGAAAAGCAATATGACGAGATAT 116
Qy 72 AGTAGTATCTAAACCTTAAAGCAAGATGAGGTCATACAGCATATGGAATTGA 131
Db 117 AATAGTATCTAGAAATCAAGCAAAAGAAAGTATACAGGATTTGATGAAATTC 176
Qy 132 ATATGACTAAGAAAATATATTTAGCTTACGCTTGGCTGTAAGTGTATATCAATAT 191
Db 177 ATATGACTAAGAAAAGTAAATATATGATATGATGCTTTACCGTAAAGTATATCAATAT 236
Qy 192 ATCTAACTTCCGTGATGATATTAAGATTAAGAGACTGCTGAGAAATTTTCAAGA 251
Db 237 CTCTAACCTTCCAGAGACTATATTAATAATAGAGATTAACAGGTATATTTCAAGA 296
Qy 252 ATATATGCTATATCTTCACTGCACTATATAGTGAATGATGATTAATTTCTAAGA 311
Db 297 ATACATGCTATATCTTATCTGCACTATATGAGTGAATGATGATTAATTTCTAAGA 356
Qy 312 TATGCAAAATGCTTTTAAAGATGAACGATTTTGAAGTCAATATCTCAAAACAT 371
Db 357 TATGCAAAATGATTTTAAACCAAGAACTGAGATTTGAGGTCAACCTTTAAATAAT 416
Qy 417 TTGAAATGCTGCTGAGCTTGAATAATTAACATGAGTCTTATTCAGATGAGATTAAT 431
Db 477 CTGGATATTTCTGATCTTGAATTAACATTTGGAAGATTTATGATGAAATTAAT 476
Qy 432 ATTAGCACTATATTTTCTGTAACAGAAATTCACCTGAGAGAAATCAACATCAAA 491
Db 477 ATTAGCACTATATTTCTTGTCTCAAGAACTTCAATGAGGCAATCAACATCAAA 536
Qy 492 TGCCCAAGATTTTAAATTAATGATTTCTTATTTACCTTATGCTGTAATCTGACT 551
Db 537 TGCCCAAGATTTTAAATTAATGATTTTCTTATTTATCTTATGCTGTAATCTGACT 596
Qy 552 GGGAGAGAGATTTTCAAAAACCTTTAACAATGATTAAGGCTAATATCAATGAGAA 611
Db 597 GGGAGAGAGATTTTCAAAAACCTTTAACAATGATTTAACAATGATTTAAGAGAA 656
Qy 612 TTATATGAGAGAAAAAATCTTTTAAACCTTTCTTGAACCAACGAGAGATTAACCTGA 671
Db 657 TTATATGAGAGAAAAAATCTTTTAAACCTTTCTTGAACCAACGAGAGATTAACCTGA 716
Qy 672 TGCGAATATGATTTTGGCTGACCAACAGAGCGCTTAA----- 714
Db 717 TGCGAAGAACGCTACTTGGCGGCTCAACAAAGCGCTTAAATTGCCAACACGCTTTC 776
Qy 715 -----TGAGAGTGAAGTTTAAAGACTTAA 740
Db 777 TACAGCAACAAAGCTTACAGACAGCTTATATTTGAGAGATTTGCAAAAACCTTAG 836
```

```

Qy 741 AATAACAAATCTAGAAATGATTTTCTAATATGGAAGGGCTGCAGAAAACAAAGTATAG 800
Db 837 AGATTAACCATTCAGAAATATCATTTATATAAATGATATGCTGCAAAAACGAAATATAG 896
Qy 801 TTCAATTTTAAAGAGGTACAAAAGGGTACGCTCCACAGACAGAGGAAAGTATTGG 860
Db 897 TTCAATTTTAAAGAGGTACAAAAGGGTATGATCATGTCAGACAGAGCAATATTGG 956
Qy 861 TACAGCCAGTGGCAATACCTGCAAAAATTCGCCAATA---ATTATATAGTGTAGGCT 917
Db 957 TACAAAAGGCGCGAGTAACTTCGAAAACCTGCAAGGTAGAGATTATATAGATTAAGCT 1016
Qy 918 AAGCCAAAAGACAGGGTACCTTACTCAAAATGATCTGCAATATACATGACGGTTCA 977
Db 1017 AAGCCAAAGACAGGGTAACTTCTCCATTAATATATCTGACCAATATATGAGATCCA 1076
Qy 978 TAGGTGGAACTCATTTATTAATAATATATGA 1008
Db 1077 AAGTGTGGAACCTCATTTACCAAAATATATATA 1107

```

RESULT 3

```

US-10-617-962-5
; Sequence 5, Application US/10617962
; Publication No. US20040055036A1
GENERAL INFORMATION:
; APPLICANT: Commonwealth Scientific and Industrial Research Organisation
; APPLICANT: East, Peter David
; TITLE OF INVENTION: Toxin Genes from the Bacteria Xenorhabdus nematophilus and Photox
; TITLE OF INVENTION: Luminescens
; FILE REFERENCE: 050179-0076
; CURRENT APPLICATION NUMBER: US/10/617,962
; CURRENT FILING DATE: 2003-07-14
; PRIOR APPLICATION NUMBER: US/09/463,048A
; PRIOR FILING DATE: 2002-12-13
; PRIOR APPLICATION NUMBER: PCT/AU98/00562
; PRIOR FILING DATE: 1998-07-17
; PRIOR APPLICATION NUMBER: PO 8088
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: Patent version 3.1
; SEQ ID NO 5
; LENGTH: 1205
; TYPE: DNA
; ORGANISM: Xenorhabdus nematophilus
US-10-617-962-5

```

```

Query Match          65.9%; Score 663.8; DB 16; Length 1205;
Best Local Similarity 79.4%; Pred. No. 2,2e-127;
Matches 835; Conservative 0; Mismatches 162; Indels 54; Gaps 2;

```

```

Qy 12 AATTAACACCTGATAGTAGAGTATATCCACCCGTTGAAAAGCAATATGACGAGATAT 71
Db 73 AGTAACGCGCTGATGATTAAGGTGAATATCAACCCGTTGAAAAGCAATATGCGGAGATAT 132
Qy 72 AGTATGATATCTAACTTTAAGCAACAGATGAGGTCATACAGCATCTATATGAAATGA 131
Db 133 AATAGGTGATCTAGAAATTCACAAACAAATGAAGTCTATACAGGATTTGATGAAATGC 192
Qy 132 AATAGGAGTAAAGAAATATATATTTAGCTTACGCTTGGCTGATAGGTATCATATGT 191
Db 193 AATAGGAGTAAAGAAATATATATTTAGCTTACGCTTGGCTGATAGGTATCATATGT 252
Qy 192 ATCTAAACTCTCTGATGATCTATATATAGAAATTAAGAGACTGCTGAGAGATTTATCAGA 251
Db 253 CTCTCAAACTTCCAGAACTATATATTAATAAGATTAACAAGGATGAAATTTATCAGAA 312
Qy 252 AATAATGCTATATCTTCACTGCACTATAGTGAATGGAATTCATTAAGA 311
Db 313 AATAATGCTATATCTTCACTGCACTATAGTGAATGGAATTCATTAAGA 372
Qy 312 TATGGCAAAATGATTTTATATAGAAATGAACTGATTTTGAAGGTCAATATCTCAAAACAT 371

```

```

Db 373 TATGGCAAAATGATTTTATCCAGAAAGAACTGAGCTTTGAGAGTCAACCTTTAAATAAC 432
Qy 372 TTGAATGTTCTGAGGCTGAAAATTAACATGAGTGCTTATTCAGATGACGATTAAT 431
Db 433 CTGGAAATATCTCATCTTGGAATTAACATTTGGAAGATTTATCAGATGAAGTAAAT 492
Qy 432 ATTAGCACTATATTTTCTGTACAGAAATTCACATGAGGAAATTCACAAATCAA 491
Db 493 ATTAGCACTATATTTCTTGTGTACAGAAATTCATGAGGAAATTCACAAATCAA 552
Qy 492 TGCCGCAAGATTTTAAATTAATGATTTCTTATTTACCTTATCTGTCTGATCTCT 551
Db 553 TGCGCAAAATTTTAAAGTAAATGATTTTCTTATTTATCTTATCTGTCTGATCTCT 612
Qy 552 GGGAAAGAGATTTTTCACAAAACCTTACAAATGATTAAGGCTAAATCATTAAGAAA 611
Db 613 GGGAAAGAGATTTTTCACAAAATTTTACAAATGATTTACAAATCATTAATCATTAAGAAA 672
Qy 612 TTATATTTGAGAGAAAATTTCTTAAACCTTTCTTTCGACCAACGCGAGATTAACCTGA 671
Db 673 TTATATTTGAGAGAAAATTTCTTAAACCTTTCTTTCGACCAACGCGAGATTAACCTGA 732
Qy 672 TGCGCAATAGGATTTTGTGCTGACCAACGAGCGCTTAA----- 714
Db 733 TGCGCAATAGGATTTTGTGCTGACCAACGAGCGCTTAAATTTGCGCAACAGCTCTTC 792
Qy 715 -----TGAGAGTGAAGTTTAAAGACTTAA 740
Db 793 TACAGCAACAACTGCTACAGAGCTTCACTATTTGAGAGTTATTTGCAAAAACCTTAC 852
Qy 741 AATAACAAATCTAGAAATGATTTTCTAATATGGAAGGGCTGCAGAAAACAAAGTATAG 800
Db 853 AGATTAACCATTCAGAAATATCATTTATATAAATGATATGCTGCAAAAACGAAATATAG 912
Qy 801 TTCAATTTTAAAGAGGTACAAAAGGGTACGCTCCACAGACAGAGGAAAGTATTGG 860
Db 913 TTCAATTTTAAAGAGGTACAAAAGGGTATGATCTCAGTACAGAGCAAGATTTGG 972
Qy 861 TACAGCCAGTGGCAATACCTGCAAAAATTCGCCAATA---ATTATATAGTGTAGGCT 917
Db 973 TACAAAAGGCGCGAGTAACTTCGAAAACCTGCAAGGTAGAGTTATATATGATTAAGCT 1032
Qy 918 AAGCCAAAAGACAGGGTAACTTACTCAAAATGATATCTGCAATATACATGACGGTTCA 977
Db 1033 AAGCCAAAGACAGGGTAACTTCTCCATTAATATATCTGACCAATATATGAGATCCA 1092
Qy 978 TAGGTGGAACTCATTTATTAATAATATATGA 1008
Db 1093 AAGTGTGGAACCTCATTTACCAAAATATATATA 1123

```

RESULT 4

```

US-10-235-192A-27
; Sequence 27, Application US/10235192A
; Publication No. US20040043389A1
GENERAL INFORMATION:
; APPLICANT: McCarthy, Jeanette
; TITLE OF INVENTION: Methods and Compositions for Identifying
; TITLE OF INVENTION: Risk Factors for Abnormal Lipid Levels and the Diseases
; FILE REFERENCE: MMI-011
; CURRENT APPLICATION NUMBER: US/10/235,192A
; CURRENT FILING DATE: 2002-09-04
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 27
; LENGTH: 188971
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 145329-145428, 187896-187995, 195894-195993
; OTHER INFORMATION: N = any nucleotide

```

US-10-235-192A-27

Query Match 5.0%; Score 50.6; DB 16; Length 188971;
Best Local Similarity 42.6%; Pred. No. 9.3;
Matches 263; Conservative 0; Mismatches 354; Indels 0; Gaps 0;

QY 12 ATTAACCTGATGATGAGAGTATCCACCCCTGGAAGCAATGCGAGATTT 71
DB 103629 ATCTATATATGAAATATATATGAAATATCTATACAGAAATCTATATGAAATATCTATAT 103688
QY 72 AGTAGCTATCTAAACCTTAAGCAACAGATGAGGCTCATACAGCATCATGAGATTGA 131
DB 103689 ATGAAATATCTATATGAAATATCTATATGAAATATCTATATGAAATATCTATATGAAATATCT 103748
QY 132 ATATGAGCTAAGAAATATATATATGCTTACCTTGGCTGTAGTGTATTCATATGT 191
DB 103749 ATATGAAATATCATGAAATATATATGAAATATATGAAATATCTATATGAAATATAT 103808
QY 192 ATCTAACTCTCTGATGATATATTAAGAAATTAAGAGACTGCTGAGAGAAATTTATCAGA 251
DB 103809 ATGAATATCTATGAAATATCTATATGAAATATATGAAATATCTATATGAAATATATATG 103868
QY 252 ATATATGCTAACTCTCTGATGATATTAAGAAATTAAGAGACTGCTGAGAGAAATTTATCAGA 311
DB 103869 AATTTATATGAAATTTATATATGAAATATATATGAAATATATATGAAATTTATATGAA 103928
QY 312 TATGCAAAATGCTTTTATATAGAAATGAACTGATTTTGAAGTCAATATCTCAAAAACAT 371
DB 103929 TATATATATGAAATATATATATGAAATATATATGAAATATATATGAAATATATATGAAATA 103988
QY 372 TTGGAATGTTCTGAGCTTGAAATTAACCATTTGAGTCTTATTCAGATGAGATTAAT 431
DB 103989 TATATATGAAATATATATATGAAATATATATGAAATATATATGAAATATATATGAAATATA 104048
QY 432 ATTACACATATATTTTCTCTGTCAGAGAAATTCACATGAGAAATCAACAATCANA 491
DB 104049 TATATGAAATATATATGAAATATATATGAAATATATATGAAATATATATGAAATATATATA 104108
QY 492 TGCCGCAAGATTTTAAATTAATGATTTCTTAATTTACCTTATCTGCTGTAACCTGACT 551
DB 104109 TGAATATATATGAAATATATATATGAAATATATATGAAATATATATGAAATATATATGAA 104168
QY 552 GCGAAGAGATTTTTCAAAACTTTTACATGATTAAGCGCAATCTTGAAGA 611
DB 104169 ATTATATATGAAATATATATGAAATATATATGAAATATATATGAAATATATATGAAAT 104228
QY 612 TTATATGAGAGAAAA 628
DB 104229 ATATATGAAATATATATA 104245

RESULT 5
US-10-240-453-95
; Sequence 95, Application US/10240453
; Publication No. US20030148326A1

GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA
; TITLE OF INVENTION: Transcription
; TITLE OF INVENTION: by Means of Assessing the Methylation Status of Genes Associated
; FILE REFERENCE: 5013.1009
; CURRENT APPLICATION NUMBER: US/10/240,453
; CURRENT FILING DATE: 2002-10-02
; PRIOR APPLICATION NUMBER: PCT/EP01/03973
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7

;; PRIOR FILING DATE: 2000-06-30
;; PRIOR APPLICATION NUMBER: DE 10043826.1
;; PRIOR FILING DATE: 2000-09-01
;; NUMBER OF SEQ ID NOS: 350
;; SEQ ID NO 95
;; LENGTH: 11394

;; TYPE: DNA
;; ORGANISM: Artificial Sequence
;; FEATURE:
;; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-240-453-95

Query Match 4.9%; Score 49.8; DB 15; Length 11394;
Best Local Similarity 50.6%; Pred. No. 4.8;
Matches 120; Conservative 0; Mismatches 117; Indels 0; Gaps 0;

QY 124 GGAATGGAATTCGAGCTAGAAATATATATAGCTTGAGCTTGAGTGTAT 183
DB 318 GGAATGGAATTCGAGCTAGAAATATATATAGCTTGAGCTTGAGTGTAT 377
QY 184 CATATGATCTAACTCTGATGATCTTATATAGAAATTAAGAGACTGCTGAGAGAAAT 243
DB 378 AT 437
QY 244 TATCAAGATATATATCTTATCTTATCTGACATCTGACATTTAGTGAAGAGTGAATAT 303
DB 438 AAGGTTGTATTTAT 497
QY 304 TCTAAGATATGCAAAATGTTTATATAGAAATGAACTGAAATTTGAAGTCAATAT 360
DB 498 TGAATTTTATATGAAATGATTTAT 554

RESULT 6
US-10-349-680-73
; Sequence 73, Application US/10349680
; Publication No. US20030176654A1

GENERAL INFORMATION:
; APPLICANT: Casseil, Gail
; APPLICANT: Chen, Ellison
; APPLICANT: Glass, Jennifer
; APPLICANT: Glass, John
; APPLICANT: Heiner, Cheryl
; APPLICANT: Lefkowitz, Elliott
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND METHOD FOR DETECTING UREAPLASMA
; TITLE OF INVENTION: UREALYTICUM
; FILE REFERENCE: UAB-13403/22
; CURRENT APPLICATION NUMBER: US/10/349,680
; CURRENT FILING DATE: 2003-01-23
; PRIOR APPLICATION NUMBER: US 09/601,198
; PRIOR FILING DATE: 2000-12-08
; PRIOR APPLICATION NUMBER: PCT/US99/01972
; PRIOR FILING DATE: 1999-01-29
; PRIOR APPLICATION NUMBER: US 60/073,189
; PRIOR FILING DATE: 1998-01-30
; NUMBER OF SEQ ID NOS: 181
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 73
; LENGTH: 1866
; TYPE: DNA
; ORGANISM: Ureaplasma urealyticum
US-10-349-680-73

Query Match 4.8%; Score 48.8; DB 15; Length 1866;
Best Local Similarity 47.3%; Pred. No. 4;
Matches 249; Conservative 0; Mismatches 267; Indels 10; Gaps 3;

QY 122 ATGAATGGAATATGAGCTAGAAATATATATAGCTTGAAGCTTGAGTGTAT 181
DB 1037 ATGAATGGAATATGAGCTAGAAATATATATAGCTTGAAGCTTGAGTGTAT 1096
QY 182 TTCAATATGATCTTAACTCTGATGATCTTATATAGAAATTAAGAGCTGCTGAGAA- 240

Db 1097 TTCAAAAATATATGATCATATATTTAGATATTAATTTAATAAATTTCAAGAGCAT 1156
Qy 241 --ATTATCAAGAAATATATGCTATCTTCATCTGCACATTTAGTGAAATGGTGATC 298
Db 1157 ATATTACGAAAAATTAATTAATAATCTTTTATGATGCTTTCATATATATGATGTC 1216
Qy 239 AAATTTCTAAGATATGCGAAATGGTTTTTATAGAAATGAACGTGATTTTGAAGTCAT 358
Db 1217 AAATTACAAAATATCTAATTTGACAAAATGATAGAAATACCAAAATTAATTAATTAAC 1276
Qy 359 ATCTCAAAACATTTGGATGTTCTGAGCTTGAAAATTAACCATTTGAGTGTATTGAG 418
Db 1277 AATTAATGAATTTAGTATATTTTAAATCCAAATACCAAGATCTTAATGAT 1336
Qy 419 ATGAGCAATTAATTTTATTTTCTGTCGTAGGAAATTCACCTGGA 473
Db 1337 ATAGCAAAATTAATGATTTTGAATAATTTTTCACAAATGGTATTTTATTTAA 1396
Qy 474 GGAATATCAACATCAATGCGCAAGATTTTAAATTAATGATTTCTTATTTACCTT 533
Db 1397 TGAATTTAAGCA--CTTGAATTAAGTTTAAATTTGATTTTACCTTACCAAAACAAT 1454
Qy 534 ATCTGCTGTAACTTCACTGGGAGAGGATTTTTCAAAAAATTTTCAATGATTTAGA 593
Db 1455 AAATCTTATCTAAGAGATGTTAGACCAAGTGAATAATTTTAAAGGTTAAAAA 1514
Qy 594 GGCTAAATCATTTAGGATTTATTTAGAGAAAAAATTTCTTAA 639
Db 1515 ACTTAACCATATAGAGATTTATTTCTTAAATTAATTTAGAA 1560

RESULT 7

US-10-301-533-23
; Sequence 23, Application US/10301533
; Publication No. US20030203473A1
; GENERAL INFORMATION:
; APPLICANT: Godzik, Adam
; APPLICANT: Reed, John C.
; TITLE OF INVENTION: Microbial Sumo Protease Homologs
; FILE REFERENCE: P-LI 5488
; CURRENT APPLICATION NUMBER: US/10/301,533
; CURRENT FILING DATE: 2002-11-20
; PRIOR APPLICATION NUMBER: US 60/331,895
; PRIOR FILING DATE: 2001-11-20
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 23
; LENGTH: 2767
; TYPE: DNA
; ORGANISM: Plasmodium falciparum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (2) ... (1507)
US-10-301-533-23

Query Match 4.8%; Score 48.2; DB 15; Length 2767;
Best Local Similarity 56.7%; Pred. No. 6.1;
Matches 89; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

Qy 243 TTATCAAGATATATGCTAATCTTTCATCTGCACATTTAGTGAAATGGTGATCAAT 302
Db 562 TGATCTAGTAAATATCTCTGCTTTCTTAATTAATTAATGAGAAAAATGACAAAT 621
Qy 303 TTCTAAGATATGCGAAATGGTTTTTATAGAAATGAACGTGATTTTGAAGTCAATTC 362
Db 622 GTGTATAGTAAGTAAGATCATTTTAAATTAATTAATTAATTAATTAATTAATTAAT 681
Qy 363 TCAAAACATTTGAGATGCTGAGCTTGAATAAA 399
Db 682 TGATGAGTATACATTAATATAGATGATCTATTA 718

RESULT 8

US-10-311-455-597
; Sequence 597, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEBENROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
; TITLE OF INVENTION: cycosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 597
; LENGTH: 16217
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 12269
; OTHER INFORMATION: n is a or g or c or t
US-10-311-455-597

Query Match 4.7%; Score 47.6; DB 15; Length 16217;

Best Local Similarity 43.3%; Pred. No. 16;
Matches 221; Conservative 0; Mismatches 289; Indels 0; Gaps 0;

Qy 124 GGAATGGAATATCGAGCTAAGAAATATATTTAGCTTACGCTTGGCTGATGATTT 183
Db 2495 GAAATATATGATTTTATTAATAATAATAATAATAATAATAATAATAATAATAATAATA 2554
Qy 184 CATATGATCTAAATCTCGATGACATTTATAGAAATTAAGACATCTGAGGATTT 243
Db 2555 AAGAAAAATTAATAATATTTATTAATTTTATTAATAATAATAATAATAATAATAATA 2614
Qy 244 TATCAAGATATATGCTAATCTTTCATCTGCACATTTAGTGAAATGGTGATCAATTT 303
Db 2615 TGGGTTGTTTATGTTTATATTTTATTTATTTATTTATTTATTTATTTATTTATTTAT 2674
Qy 304 TCTAAGATATGCGAAATGGTTTTTATAGAAATGAACGTGATTTGAAGTCAATATCT 363
Db 2675 TATATTATATGAAATTAATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 2734
Qy 364 CAAAACATTTGGAATGTTCTCGAGCTTGAATAATAACCATTTGAGTGTATTTAGATGAC 423
Db 2735 AAGGTAAATTAATTTTGGTATTTGAATTTATTTATTTATTTATTTATTTATTTATTT 2794
Qy 424 GATAATTTATAGACATTAATTTTCTGTCGACGAAATTTCCATGAGAAATGCA 483
Db 2795 AGTATTTAGTTAGATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 2854
Qy 484 CAATCAATGCGCAAGATTTTAAATTAATTTATTTATTTATTTATTTATTTATTTATTTAT 543
Db 2855 TTTTATTTATTTAGAAATATTTGTTTGTGTGTTAGTTTATTTATTTATTTATTTATTT 2914
Qy 544 ACTTCACTGGAGAGGATTTTTCACAAAATTTTCAATGATTTGAGGCTTAATCA 603
Db 2915 AGTATGATTTATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 2974
Qy 604 TTAGGATTTATTTAGAGAAAAAATCTT 633
Db 2975 AGTGGATTTTATTTAGAGAAATTTTAAATTT 3004

RESULT 9

US-08-781-986A-531
; Sequence 531, Application US/08781986A
; Publication No. US20030054436A1
; GENERAL INFORMATION:
; APPLICANT: Charles Kunesh
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 5255
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/781,986A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Benson, Bob
; REGISTRATION NUMBER: 30,446
; REFERENCE/DOCKET NUMBER: PB248PP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 531:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 942 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-781-986A-531

Query Match 4.7%; Score 47.4; DB 8; Length 942;
Best Local Similarity 47.8%; Pred. No. 6;
Matches 138; Conservative 0; Mismatches 151; Indels 0; Gaps 0;

QY 66 AGATATAGTACCTAATCTTAAGCAACAGATGAGGTCATACAGCATCATATG 125
DB 313 AGAAAACGTACTTAAGCAATGAGCAATTAAGTAAGCAATTAAGTCACTTCTAT 372
QY 126 AATTGAATATGAGCTAAGAAAATATATTAAGCTTAAGCTTGGCTGTAAGGTATTC 185
DB 373 AGATGAATATTAAGAGCTTAAGCTAAGTATGATATGATGATGATGATGATGATG 432
QY 186 TAATGATATTAAGCTTCCGATGACATATATAAGAAATAAGACCTGCGAGAAATTA 245
DB 433 CATAGAGAGAAATATATATGATGATGATGATGATGATGATGATGATGATGATGATG 492
QY 246 TCAAGAAATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 305
DB 493 AACGCAATCAATGCTGATTTTGTATTTTATGATATGATGATGATGATGATGATGATG 552
QY 306 TAAAGATATGCAAAATGTTTATTAAGATGAATGAACTGATTTTGAAGT 354
DB 553 AAAACAGAGCAATACGCTTTTATTAAGAAAGAAATCTATTAAGAT 601

RESULT 10
US-10-329-624-531
; Sequence 531, Application US/10329624
; Publication No. US20040043037A1
; GENERAL INFORMATION:
; APPLICANT: Charles Kunesh
; GIL H. Choi

Patrick S. Dillon
; Craig A. Rosen
; Steven C. Barash
; Michael R. Fannon
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 5256
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/329,624
; FILING DATE: 27-Dec-2002
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/956,171
; FILING DATE: October 20, 1997
; APPLICATION NUMBER: 60/009,861
; FILING DATE: January 5, 1996
; APPLICATION NUMBER: 08/781,986
; FILING DATE: January 3, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Mark J. Hyman
; REGISTRATION NUMBER: 46,789
; REFERENCE/DOCKET NUMBER: PB248P1D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (240) 314-1224
; TELEFAX: (301) 309-8439
; INFORMATION FOR SEQ ID NO: 531:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 942 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 531:
US-10-329-624-531

Query Match 4.7%; Score 47.4; DB 16; Length 942;
Best Local Similarity 47.8%; Pred. No. 6;
Matches 138; Conservative 0; Mismatches 151; Indels 0; Gaps 0;

QY 66 AGATATAGTACCTAATCTTAAGCAACAGATGAGGTCATACAGCATCATATG 125
DB 313 AGAAAACGTACTTAAGCAATGAGCAATTAAGTAAGCAATTAAGTCACTTCTAT 372
QY 126 AATTGAATATGAGCTAAGAAAATATATTAAGCTTAAGCTTGGCTGTAAGGTATTC 185
DB 373 AGATGAATATTAAGAGCTTAAGCTAAGTATGATATGATGATGATGATGATGATG 432
QY 186 TAATGATATTAAGCTTCCGATGACATATATAAGAAATAAGACCTGCGAGAAATTA 245
DB 433 CATAGAGAGAAATATATATGATGATGATGATGATGATGATGATGATGATGATGATG 492
QY 246 TCAAGAAATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 305
DB 493 AACGCAATCAATGCTGATTTTGTATTTTATGATATGATGATGATGATGATGATGATG 552
QY 306 TAAAGATATGCAAAATGTTTATTAAGATGAATGAACTGATTTTGAAGT 354
DB 553 AAAACAGAGCAATACGCTTTTATTAAGAAAGAAATCTATTAAGAT 601

RESULT 11
US-10-211-179-11
; GENERAL INFORMATION:
; APPLICANT: Nicholas M. Dean


```
2000-06-30
2000-09-01
NUMBER OF SEQ ID NOS: 178
SEQ ID NO 68
LENGTH: 11052
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-257-166-68
```

```
Query Match 4.7%; Score 47; DB 16; Length 11052;
Best Local Similarity 48.7%; Pred. No. 18;
Matches 128; Conservative 0; Mismatches 135; Indels 0; Gaps 0;
```

```
225 AGAGCTGCTGAGAAATTTATCAAGATATGCTATCTTCATCTGCACTATTGG 284
1196 AGAAGTATAGATATTTTAAAGTTTGGGTTTTTTAGAAAGTATATATTGG 1255
285 TGAATAATGCTATCAAAATTTCTAAAGATATGCAAAATGGTTTTTATTAAGATGAACCTGA 344
1256 AGTAAATGTTTAAAGAGTGAAGAAAGGAGGTTTTTATATATGCGATTAAGTT 1315
345 TTTGAAGTCAATATCCCAAAACATTGGATGTTCTGAGCTTGAATAAACATT 404
1316 TTTATAGTATTTTATTAAGAGGGGTTGGGTGAAGGTATTAATAAATTAAGTT 1375
405 GAGTCTTATTCAGATGACGATTAATTTATGACACTATATTTTCTCTGTACAGGAAT 464
1376 AAAAATTAAATTAATTTTGTATATTTTATGTTATTTTGTGTTTATAGGATTA 1435
465 TCCACTGAGAGAAATCAACAAT 487
1436 TTGAGGTTTGAATAATTATAGT 1458
Db
```

RESULT 14

```
US-10-424-599-100851
Sequence 100851, Application US/10424599
Publication No. US20040031072A1
GENERAL INFORMATION:
APPLICANT: La Rosa Thomas J
APPLICANT: Kovalic David K
APPLICANT: Zhou Yihua
APPLICANT: Cao Yongwei
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement
CURRENT APPLICATION NUMBER: US/10/424,599
CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ ID NOS: 285684
SEQ ID NO 100851
LENGTH: 260
TYPE: DNA
ORGANISM: Glycine max
FEATURE:
OTHER INFORMATION: Clone ID: PAT_MRT3847_62081C.1
US-10-424-599-100851
```

```
Query Match 4.6%; Score 46.4; DB 16; Length 260;
Best Local Similarity 52.6%; Pred. No. 6;
Matches 101; Conservative 0; Mismatches 91; Indels 0; Gaps 0;
```

```
249 AGAATATATGCTAATCTTTCATCTGCACTATTAAGTGAATAATGGTATCAATTTCTAA 308
1 ATAACTGTTAATCTTTTAAACATTTTAAAGCAAAAAGGACTGAAGTTTCGAA 60
309 AGATATGCAATGCTTTTATTAAGATGAATGATTTTGAAGGCAATATCTCAAA 368
61 AAAAGAGAAATAGTTTCAAGTCTTCTTACCTGAAATTTTAAATCAAAAC 120
369 CATTGGAATGTTCTGACCTTGAATAATTAACCATTTGAGTGTATTCAGATGACATTA 428
```

```
Db 121 CGTTGATTTTAAAAAATGACATTAACAAAAGTTTGCTATCTTAAGGGGGGAAA 180
Qy 429 ATTATGACACT 440
Db 181 AACTTACTACT 192
```

```
RESULT 15
US-10-322-281-700
Sequence 700, Application US/10322281
Publication No. US20040126762A1
GENERAL INFORMATION:
APPLICANT: David W. Morris
APPLICANT: Marc S. Malandro
TITLE OF INVENTION: Novel Compositions and Methods in Cancer
FILE REFERENCE: 529452001000
CURRENT APPLICATION NUMBER: US/10/322,281
CURRENT FILING DATE: 2002-12-17
NUMBER OF SEQ ID NOS: 866
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 700
LENGTH: 34722
TYPE: DNA
```

```
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(34722)
OTHER INFORMATION: n = A,T,C or G
US-10-322-281-700
```

```
Query Match 4.6%; Score 46.2; DB 17; Length 34722;
Best Local Similarity 45.0%; Pred. No. 40;
Matches 174; Conservative 0; Mismatches 213; Indels 0; Gaps 0;
```

```
50 AAAAGCAATGCGAGATATAGTACTATCTTAATCTTAAAGCAACAGATGAGGTC 109
2706 AAAATTTATTAAGAGAAACAAAATTAATATATATATATATATATATATATATAT 2765
110 ATACGATCATATGATGAATGATATCGAGCTAAGAAATATATATATATATATATAT 169
2766 ATTAATTTATATATATATATATATATATATATATATATATATATATATATATAT 2825
170 CTGTAAGTGATTCATATATGATCTAATCTTCTGATGACTATATATATATATATAT 229
2826 ATGATATATATATATATATATATATATATATATATATATATATATATATATAT 2885
230 CTGCGAGAGATTTATGAGAAATATATATATATATATATATATATATATATATAT 289
2886 TATGTTTATATATATATATATATATATATATATATATATATATATATATATAT 2945
Qy 290 ATGGTATCAAAATTTCTAAAGATATGCAAAATGTTTTTATTAAGATGACGATTTTG 349
Db 2946 TTTATATATATATATATATATATATATATATATATATATATATATATATATATAT 3005
350 AAGTCAATATCTTCAAAACATTTGGAATGTTCTGACCTGAAATTAACCATTTGAGTG 409
Db 3006 AATATATATATATATATATATATATATATATATATATATATATATATATATAT 3065
Qy 410 CTTATTCAGATGACATTAATTTATG 436
Db 3066 ATTATATATATATATATATATATATATATATATATATATATATATATATATAT 3092
```

```
Search completed: November 21, 2004, 08:42:21
Job time : 589 secs
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 21, 2004, 07:28:02 ; Search time 116 Seconds
(without alignments)
6176.510 Million cell updates/sec

Title: US-10-617-962-2
Perfect score: 1008
Sequence: 1 atggttatacaatracacc.....ctcatataaataatcatga 1008

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 824507 seqs, 355394441 residues

Total number of hits satisfying chosen parameters: 1649014

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA:
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PTUS.COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1008	100.0	1008	US-09-463-048A-2	Sequence 2, Appli
2	663.8	65.9	1107	US-09-463-048A-1	Sequence 1, Appli
3	663.8	65.9	1205	US-09-463-048A-5	Sequence 5, Appli
4	628.2	62.3	1272	US-08-569-168-6	Sequence 6, Appli
5	521.4	51.7	834	US-08-569-168-3	Sequence 3, Appli
6	521.4	51.7	837	US-08-569-168-1	Sequence 1, Appli
7	62.6	6.2	1141	US-08-806-708B-22	Sequence 22, Appli
8	59.8	5.9	7218	US-08-232-463-14	Sequence 14, Appli
9	51	5.1	1055	US-09-806-708B-23	Sequence 23, Appli
10	50.6	5.0	1141	US-09-806-708B-22	Sequence 22, Appli
11	48.8	4.8	1866	US-09-601-198-153	Sequence 153, App
12	47.8	4.7	729	US-09-134-001C-1161	Sequence 1161, Ap
13	47.6	4.7	1055	US-09-806-708B-23	Sequence 23, Appli
14	47.4	4.7	942	US-08-956-171E-531	Sequence 531, App
15	47.4	4.7	942	US-08-781-966A-531	Sequence 531, App
16	47.4	4.7	41708	US-09-470-512A-3	Sequence 3, Appli
17	47.4	4.7	41708	US-09-676-519-18	Sequence 18, Appli
18	47	4.7	5340	US-09-627-122-21	Sequence 21, Appli
19	44.8	4.4	3399	US-09-614-221A-600	Sequence 600, App
20	44.6	4.4	13508	US-08-956-171E-120	Sequence 120, App
21	44.6	4.4	13508	US-08-781-966A-120	Sequence 120, App
22	44	4.4	832	US-09-621-976-2813	Sequence 2813, App
23	43.8	4.3	665	US-08-883-795A-36	Sequence 36, Appli
24	43.2	4.3	1470	US-09-830-217-1	Sequence 1, Appli
25	42.8	4.2	7425	US-09-453-702B-212	Sequence 212, App
26	42.6	4.2	5340	US-09-627-122-21	Sequence 21, Appli
27	42.6	4.2	640681	US-09-790-988-1	Sequence 1, Appli

28	42.4	4.2	627	US-09-248-796A-14060	Sequence 14060, A
29	42.2	4.2	5136	US-08-956-171E-332	Sequence 332, App
30	42.2	4.2	5136	US-08-781-966A-332	Sequence 332, App
31	41.8	4.1	2557	US-09-710-279-3829	Sequence 3829, Ap
32	41.8	4.1	3146	US-09-710-279-4281	Sequence 4281, Ap
33	41.8	4.1	3646	US-09-710-279-4070	Sequence 4070, Ap
34	41.4	4.1	978	US-09-248-796A-5337	Sequence 5337, Ap
35	41.4	4.1	2424	US-08-956-171E-392	Sequence 392, App
36	41.4	4.1	2424	US-08-781-966A-392	Sequence 392, App
37	41.2	4.1	2142	US-09-107-532A-905	Sequence 905, App
38	41.2	4.1	8920	US-08-446-855A-1	Sequence 1, Appli
39	41.2	4.1	8920	US-09-150-741-1	Sequence 1, Appli
40	41.2	4.1	14066	US-09-601-198-56	Sequence 56, Appli
41	41	4.1	1578	US-09-248-796A-5803	Sequence 5803, Ap
42	41	4.1	43095	US-09-676-519-17	Sequence 17, Appli
43	40.8	4.0	832	US-09-621-976-2813	Sequence 2813, App
44	40.8	4.0	1830121	US-09-557-884-1	Sequence 1, Appli
45	40.8	4.0	1830121	US-09-643-990A-1	Sequence 1, Appli

ALIGNMENTS

```

RESULT 1
US-09-463-048A-2
; Sequence 2, Application US/09463048A
; Patent No. 6630619
; GENERAL INFORMATION:
; APPLICANT: Commonwealth Scientific and Industrial Research Organisation
; TITLE OF INVENTION: Toxin Genes from the Bacteria Xenorhabdus nematophilus and Photor
; FILE REFERENCE: 050179-0076
; CURRENT APPLICATION NUMBER: US/09/463,048A
; PRIOR FILING DATE: 2002-12-13
; PRIOR APPLICATION NUMBER: PCT/AU98/00562
; PRIOR FILING DATE: 1998-07-17
; PRIOR APPLICATION NUMBER: PO 8088
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 1008
; TYPE: DNA
; ORGANISM: Photorhabdus luminescens
US-09-463-048A-2

Query Match      100.0%; Score 1008; DB 4; Length 1008;
Best Local Similarity 100.0%; Pred. No. 1.2e-242;
Matches 1008; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 ATGCTTATCAATTACACCTGATGATGAGTATTCACCCGTTGAAAGCAATA 60
        1 ATGCTTATCAATTACACCTGATGATGAGTATTCACCCGTTGAAAGCAATA 60
DB      1 ATGCTTATCAATTACACCTGATGATGAGTATTCACCCGTTGAAAGCAATA 60
        1 ATGCTTATCAATTACACCTGATGATGAGTATTCACCCGTTGAAAGCAATA 60
QY      61 GCAGAGATATAGTACGATCTAACTTAAGCAACAGATGAGGTCATACAGCATCA 120
        61 GCAGAGATATAGTACGATCTAACTTAAGCAACAGATGAGGTCATACAGCATCA 120
DB      61 GCAGAGATATAGTACGATCTAACTTAAGCAACAGATGAGGTCATACAGCATCA 120
        61 GCAGAGATATAGTACGATCTAACTTAAGCAACAGATGAGGTCATACAGCATCA 120
QY      121 TATGAATTAATGATCGAGTAAAGAAATATATTAAGCTTACGTTGGCTGAAGTGT 180
        121 TATGAATTAATGATCGAGTAAAGAAATATATTAAGCTTACGTTGGCTGAAGTGT 180
DB      121 TATGAATTAATGATCGAGTAAAGAAATATATTAAGCTTACGTTGGCTGAAGTGT 180
        121 TATGAATTAATGATCGAGTAAAGAAATATATTAAGCTTACGTTGGCTGAAGTGT 180
QY      181 ATTCAATTAATGATCGAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 240
        181 ATTCAATTAATGATCGAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 240
DB      181 ATTCAATTAATGATCGAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 240
        181 ATTCAATTAATGATCGAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 240
QY      241 ATTATCAAGATATATGATCTTATCATCTGACATTAATAGTGAATAGTGAATCA 300
        241 ATTATCAAGATATATGATCTTATCATCTGACATTAATAGTGAATAGTGAATCA 300
DB      241 ATTATCAAGATATATGATCTTATCATCTGACATTAATAGTGAATAGTGAATCA 300
        241 ATTATCAAGATATATGATCTTATCATCTGACATTAATAGTGAATAGTGAATCA 300
QY      301 ATTCTTAAGATATGAGCAATGTTTATTAAGATGAATGATTTGAAAGTCAATAT 360
        301 ATTCTTAAGATATGAGCAATGTTTATTAAGATGAATGATTTGAAAGTCAATAT 360

```

Db 301 ATTCTAAAGATATGGCAATGCTTTTATTAAGATGAACGTGATTTTGAAGTCAATAT 360
 Qy 361 CCTCAAAACATTGGAAATGCTTCTGAGCTTGAATAAACAATGAGTCTTATTCAGAT 420
 Db 361 CCTCAAAACATTGGAAATGCTTCTGAGCTTGAATAAACAATGAGTCTTATTCAGAT 420
 Qy 421 GACGTAATATTTTGGCACTATATTTTCTCTGACAGAAATTCACCTGAGGAAAT 480
 Db 421 GACGTAATATTTTGGCACTATATTTTCTCTGACAGAAATTCACCTGAGGAAAT 480
 Qy 481 CAACATCAAAATGCGCAAGATTTTAAATTAATGATTTCTTTTACCTTATCTGCT 540
 Db 481 CAACATCAAAATGCGCAAGATTTTAAATTAATGATTTCTTTTACCTTATCTGCT 540
 Qy 541 GTAACTTCACTGGAGAGAGATTTTTCACAAAACCTTTTACAAATGATTAAGGCTAAA 600
 Db 541 GTAACTTCACTGGAGAGAGATTTTTCACAAAACCTTTTACAAATGATTAAGGCTAAA 600
 Qy 601 TCATTAAGAAATTAATTTGAGAGAAAACCTTTCTTAAACCTTTCTTTCGACACCGCAG 660
 Db 601 TCATTAAGAAATTAATTTGAGAGAAAACCTTTCTTAAACCTTTCTTTCGACACCGCAG 660
 Qy 661 AGATTACCTGATGGAGAAATAGTATTTTGGCTGACCAAGAGCGCTTAATGAGAGA 720
 Db 661 AGATTACCTGATGGAGAAATAGTATTTTGGCTGACCAAGAGCGCTTAATGAGAGA 720
 Qy 721 GTGACTTTTAAAGAACTTAAATAAACAATCTAGAAATGATTTTCTAATATGAAAGG 780
 Db 721 GTGACTTTTAAAGAACTTAAATAAACAATCTAGAAATGATTTTCTAATATGAAAGG 780
 Qy 781 GGTGAAAACAAATATATGCTCATTTATTAAGAGGTCAAAAGGTAAAGCTCTCACAG 840
 Db 781 GGTGAAAACAAATATATGCTCATTTATTAAGAGGTCAAAAGGTAAAGCTCTCACAG 840
 Qy 841 ACAGCAGGAAAGATTTGCTACAGCAGTGGCAAGTAACTGAAAATTTGCGAATAT 900
 Db 841 ACAGCAGGAAAGATTTGCTACAGCAGTGGCAAGTAACTGAAAATTTGCGAATAT 900
 Qy 901 TTATATAGTGTAGGCTTAAGCCAAAAGACAGGGTAACTTTACTCAAAATGATACTGAC 960
 Db 901 TTATATAGTGTAGGCTTAAGCCAAAAGACAGGGTAACTTTACTCAAAATGATACTGAC 960
 Qy 961 AATAAATGACGGTTCATATGCTTGAACCTCATTTAATAAATATATGA 1008
 Db 961 AATAAATGACGGTTCATATGCTTGAACCTCATTTAATAAATATATGA 1008
 RESULT 2
 US-09-463-048A-1
 ; Sequence 1, Application US/09463048A
 ; Patent No. 6630619
 ; GENERAL INFORMATION:
 ; APPLICANT: Commonwealth Scientific and Industrial Research Organisation
 ; APPLICANT: EAST, Peter David
 ; TITLE OF INVENTION: Toxin Genes from the Bacteria *Xenorhabdus nematophilus* and *Photobacterium*
 ; FILE REFERENCE: 050179-0076
 ; CURRENT APPLICATION NUMBER: US/09/463,048A
 ; CURRENT FILING DATE: 2002-12-13
 ; PRIOR APPLICATION NUMBER: PCT/AU98/00562
 ; PRIOR FILING DATE: 1998-07-17
 ; PRIOR APPLICATION NUMBER: PO 8088
 ; NUMBER OF SEQ ID NOS: 20
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO: 1
 ; LENGTH: 1107
 ; TYPE: DNA
 ; ORGANISM: *Xenorhabdus nematophilus*
 US-09-463-048A-1
 Query Match 65.9%; Score 663.8; DB 4; Length 1107;
 Best Local Similarity 79.4%; Pred. No. 1.3e-156;

	Matches	835; Conservative	0; Mismatches	162; Indels	54; Gaps	2;
Qy 12	ATTAACTGATGATAGAAATGATATCAACCGCTTGAAGCAAAATAGAGAGATAT	71				
Db 57	AGTAACTGATGATAGAAATGATATCAACCGCTTGAAGCAAAATAGAGAGATAT	116				
Qy 72	AGTAACTGATGATAGAAATGATATCAACCGCTTGAAGCAAAATAGAGAGATAT	131				
Db 117	AATAGTGTACTGATAGAAATGATATCAACCGCTTGAAGCAAAATAGAGAGATAT	176				
Qy 132	ATATGAGCTTGAAGAAATATATATGCTTACGCTTTGCTGAAAGTATATCATATAT	191				
Db 177	ATATGAGCTTGAAGAAATATATATGCTTACGCTTTGCTGAAAGTATATCATATAT	236				
Qy 192	ATCTAACTTCTGATGATATATATGATTAAGAAATGATGATGATGATGATGATGAT	251				
Db 237	CTCTCACTTCTGATGATATATATGATTAAGAAATGATGATGATGATGATGATGAT	296				
Qy 252	ATATATGCTTATATCTTTCATCTGATATATGATGATGATGATGATGATGATGAT	311				
Db 297	ATATATGCTTATATCTTTCATCTGATATATGATGATGATGATGATGATGATGAT	356				
Qy 312	TATGCAAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	371				
Db 357	TATGCAAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	416				
Qy 372	TTGAAATGCTTCTGATGATGATGATGATGATGATGATGATGATGATGATGAT	431				
Db 417	CTGGATATTTCTGATGATGATGATGATGATGATGATGATGATGATGATGAT	476				
Qy 432	ATTAGCACTATATTTTCTCTGATGATGATGATGATGATGATGATGATGATGAT	491				
Db 477	ATTAGCACTATATTTTCTCTGATGATGATGATGATGATGATGATGATGATGAT	536				
Qy 492	TGCGCAAGATTTTAAATATATGATGATGATGATGATGATGATGATGATGATGAT	551				
Db 537	TGCGCAAGATTTTAAATATATGATGATGATGATGATGATGATGATGATGATGAT	596				
Qy 552	GGAAGAGAGATTTTCAAAAACCTTTTCAATATGATGATGATGATGATGATGAT	611				
Db 597	GGAAGAGAGATTTTCAAAAACCTTTTCAATATGATGATGATGATGATGATGAT	656				
Qy 612	TTATATGAG	671				
Db 657	TTATATGAG	716				
Qy 672	TGCGCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	714				
Db 717	TGCGCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	776				
Qy 715	-----TGAGAGTGAATTTTAAAGAACTTAA 740					
Db 777	TACAGCAACAGCTTACAGCAGCTTCAATATGAGAGATGATGATGATGATGATG	836				
Qy 741	AAATTAACAAATCTAGAGATGATGATGATGATGATGATGATGATGATGATGAT	800				
Db 837	AGATTAACCAATCTAGAGATGATGATGATGATGATGATGATGATGATGATGAT	896				
Qy 801	TTATATGAG	860				
Db 897	TTATATGAG	956				
Qy 861	TACAGCAGTGGCAGTAACTGAGAAATTTGCGCAATA--ATTATATAGTGTAGGCT	917				
Db 957	TACAGCAGTGGCAGTAACTGAGAAATTTGCGCAATA--ATTATATAGTGTAGGCT	1016				
Qy 918	AAGCCAAAAGACAGAGGATACCTTATCAAAATGATGATGATGATGATGATGAT	977				
Db 1017	AAGCCAAAAGACAGAGGATACCTTATCAAAATGATGATGATGATGATGATGAT	1076				
Qy 978	TAGTGTGAACTCATTAATAAATATATGA 1008					
Db 1077	AAGTGTGAACTCATTAATAAATATATGA 1107					

QY 12 ATTAACCTGATGATGAGATGATACCCCGTTGAAAAACAATGACGAGATAT 71
DB 139 AGTAACGCTGATGATGAGATGATACCCCGTTGAAAAACAATGACGAGATAT 198
QY 72 AGTACGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 131
DB 199 AATAGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 258
QY 132 ATATGAGCTAGAGAAATATATGATGATGATGATGATGATGATGATGATGAT 191
DB 259 ATATGAGCTAGAGAAATATATGATGATGATGATGATGATGATGATGATGAT 318
QY 192 ATCTAACTGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 251
DB 319 CTCTCACTGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 378
QY 252 ATATATGCTATCTCTGATGATGATGATGATGATGATGATGATGATGATGAT 311
DB 379 ATACATGCTATCTCTGATGATGATGATGATGATGATGATGATGATGATGAT 438
QY 312 TATGCAATGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 371
DB 439 TATGCAATGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 497
QY 372 TTGATATGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 430
DB 498 CTGGATATCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 557
QY 431 TATGCACTATCTCTGATGATGATGATGATGATGATGATGATGATGATGAT 490
DB 558 TATGCACTATCTCTGATGATGATGATGATGATGATGATGATGATGATGAT 617
QY 491 ATGCGCAATGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 550
DB 618 ATGCGCAATGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 677
QY 551 TGGGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 610
DB 678 TGGGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 737
QY 611 ATTATATGAT 670
DB 738 ATTATATGAT 797
QY 671 ATGCGCAATGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 714
DB 798 ATGCGCAATGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 857
QY 715 -----TGAAGTGAATTT-TAAAGAACCTT 738
DB 858 CTACAGCAACAGCTCTACAGCAAGCTTCACTAATGAGAGATGATGATGATGAT 917
QY 739 AAAATATCAATCTACAGCAATGATTTCTAATATGAGAGAGAGAGAGAGAGAT 798
DB 918 AGAGATTAACCCATCCAGCAATGATTTGAGAGAGAGAGAGAGAGAGAGAT 977
QY 799 AGTATATTTATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 858
DB 978 AGTATATTTATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1037
QY 859 GGTACAGCAGTGGCAGTAACTGAGAGAGAGAGAGAGAGAGAGAT 915
DB 1038 GGTACAGCAGCAGTAACTGAGAGAGAGAGAGAGAGAGAGAT 1097
QY 916 CTAAACCAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 975
DB 1098 CTAAACCAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1157
QY 976 CATAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1008
DB 1158 CAAAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1190

RESULT 5
US-08-569-168-3
; Sequence 3, Application US/08569168
; Patent No. 5972687
; GENERAL INFORMATION:
; APPLICANT: Smigielski, Adam J.
; APPLICANT: Ahnert, Raymond J.
; TITLE OF INVENTION: TOXIN GENE FROM XENORHABDUS NEMATOPHILUS
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lowe, Price, Leblanc & Becker
; STREET: 99 Canal Center Plaza, Suite 300
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22314
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/569,168
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Price, Robert L.
; REGISTRATION NUMBER: 22,685
; REFERENCE/DOCKET NUMBER: 1451-015
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-684-1111
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 834 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-569-168-3

Query Match 51.7%; Score 521.4; DB 2; Length 834;
Best Local Similarity 85.4%; Pred.No. 4.5e-121;
Matches 604; Conservative 0; Mismatches 101; Indels 2; Gaps 2;

QY 12 ATTAACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 71
DB 57 AGTAACGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 116
QY 72 AGTACGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 131
DB 117 AATAGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 176
QY 132 ATATGAGCTAGAGAAATATATGATGATGATGATGATGATGATGATGATGAT 191
DB 177 ATATGAGCTAGAGAAATATATGATGATGATGATGATGATGATGATGATGAT 236
QY 192 ATCTAACTGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 251
DB 237 CTCTCACTGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 296
QY 252 ATATATGCTATCTCTGATGATGATGATGATGATGATGATGATGATGATGAT 311
DB 297 ATACATGCTATCTCTGATGATGATGATGATGATGATGATGATGATGATGAT 356
QY 312 TATGCAATGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 371
DB 357 TATGCAATGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 415
QY 372 TTGATATGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 430
DB 416 CTGGATATCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 475


```

QY      245 ATCAAGAAATATGCTTAATCTTCAATCTGCACTATATGGTGAAGAAATGGGACCAATTT 304
Dn      774 NNVEBAATTTHTTDMCYKMTWMTWTYDMMTTHETHTTTRNMTSTNNNNNNNACTNN 715
QY      305 CTAAGAATATGGCAAAATGGTTTTTATTAAGATGAAGTGAATTTGAAGGTCAATATCTC 364
Dn      714 NNNNNNNKAYAAATANNMGCMWNNNTDARFTNNITVMBRRWMTNTKTETWYSTTRHHYGT 655
QY      365 AAAACATTTGAAATGTCCTGAGCTTGAAATAATTAACATTTGAGTCGTAT-----CAG 418
Dn      654 NNNNNNNNNNNNNNNNSCCTCTRMWTRWTKGDMTVRKVYKWRDITCTCYVDVWADSW 595
QY      419 ATGACGATTAATATTTAGACATATATTTTCTCTGAACAGAAATTCACAGAGAA 478
Dn      554 WMYANMFCRDVTTYTRNNTYCKSYAHSYTYSNNNAHYRYSAPRWSNMAKRTTRNNMM 535
QY      479 ATCAACAATCAAAATGCCGCAAGATTTTAAATTAATGATTTCTTATTTACCTTATCTG 538
Dn      534 SGBVEMRAGTWMMWHMNNNTDTRTYMMWKWARBTTTYDSCNAKSMWRGNRRAMK 475
QY      539 CTGTACTTCACTGGAGAGAGATTTTCAAAAACTTTTCAATGCAATTAGAGGCTA 598
Dn      474 MWWAANNPDAGAMDHTYMMGNNTMMRRAMKMMAMACBRAYCCNNNNNRACVWHKHKW 415
QY      599 AATCATTAAGAAATTAATATGAGAGAAAAAATCTTCTAAACCTTTCTTGCACCGC 658
Dn      414 RMTKYMMWAAACNNNNBKMYRVMAMMTSRDITNDMMWISDWBHMTVTYTMRRAM 355
QY      659 AGAGATTAACCTGATGGCAGAAATAGTTATTTGGCTGGAACCAAGAGCGCTTAATGA 718
Dn      354 NNNNNNNNRBCKTTSMMWMDHNTHTCTGNNTWGSAVBBAASWMAAGASNVYTNMCM 295
QY      719 GAGTGAATTTTAAAGAACTTAAAAATACAATCTAGAAATGCAATTTCTAATATGAAG 778
Dn      294 RMTYGGKTTNNNNNNKAMYRTKTYAACNNRRYYDPTAVTBKRNKXYCYAAYBYMYBM 235
QY      779 GGGCGCAAAAAGAAAGTATGTCATTTTAATAAAGAGTACAAAAGGGTAAAGCTCCAC 838
Dn      234 GHHHBMRRRAHRSNNMMVHCNKXKTYVSMHYHMYRBYWABAVGCGNNMKORMAHHW 175
QY      839 AGACGACAGCAAAATATTTGATACAGCAGTGGCAGTAACTGAAAAATGCGCGATA 898
Dn      174 CATNNNNMMWVAYVHHHKKKGLAATNNKTAERDHBAAVKTYYWRYDYWCAMCMNA 115
QY      899 ATTATA 905
Dn      114 KAKVATA 108

RESULT 8
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
;
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEIFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463

```

```

; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-9300
; TELEFAX: (703) 683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZspc-Fls
; US-08-232-463-14

Query Match          5.9%; Score 59.8; DB 1; Length 7218;
Best Local Similarity 6.0%; Pred.No. 2,2e-05;
Matches 25; Conservative 225; Mismatches 167; Indels 0; Gaps 0

Oy 521 TCTATTACCTGATCGCTGCTACTACCTGAGAGAGAGATTTTCAAAAACTTT 580
    |||||
Db 1484 TGTATTAACCTATCTATGCAAGTGATTAAAGATGAAGATTGTGACRRRRRRR 1422
    |||||

Oy 581 ACAATGATAGAGGCTAAATCATTAGAAATTATATTGAGAGAAAAAACTTCTAAAC 640
    :::::
Db 1424 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1365

Oy 641 CTTTCCTTCGACCGCAGAGATTACTGTAGTGCGAGAAATGTTATTGGCTGACCAA 700
    :::::
Db 1364 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1305

Oy 701 CAGAAAGCGCTAAATGAGAGTGAGTTTAAAGAACTTAAATAAACAATCTAGAAATG 760
    :::::
Db 1304 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1245

Oy 761 GATTTCATATGAGGAGGCGCTGCAAAACAAAAGTATGATTCATTTTAAAGAGTAC 820
    :::::
Db 1244 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1185

Oy 821 AAAAGGTACGCTCCACAGACAGCGAGGAAAAGTATTGTAACGCCGTGGCAGTAAC 880
    :::::
Db 1184 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1125

Oy 881 TGAAGAAATTCGCAATATATTATATAGTGAGAGCTAAGCCAAAGACAGGGGTAA 937
    :::::
Db 1124 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1068

RESULT 9
US-09-806-708B-23
; Sequence 23, Application US/09806708B
; Patent No. 6784342
; GENERAL INFORMATION:
; APPLICANT: The University of British Columbia
; TITLE OF INVENTION: Regulation of Embryonic Transcription in Plants
; FILE REFERENCE: 4810-58741
; CURRENT APPLICATION NUMBER: US/09/806,708B
; PRIOR FILING DATE: 2001-04-03
; PRIOR APPLICATION NUMBER: US 60/147,133
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 23

```



```

1  COUNTRY: USA
2  ZIP: 20850
3
4  COMPUTER READABLE FORM:
5  MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
6  COMPUTER: HP Vectra 486/33
7  OPERATING SYSTEM: MSDOS version 6.2
8  SOFTWARE: ASCII Text
9
10 CURRENT APPLICATION DATA:
11 APPLICATION NUMBER: US/08/781.986A
12
13 FILING DATE:
14
15 CLASSIFICATION: 435
16 PRIOR APPLICATION DATA:
17 APPLICATION NUMBER:
18
19 FILING DATE:
20
21 ATTORNEY/AGENT INFORMATION:
22 NAME: Benson, Bob
23
24 REGISTRATION NUMBER: 30,446
25 REFERENCE/DOCKET NUMBER: PB248PP
26 TELECOMMUNICATION INFORMATION:
27 TELEPHONE: (301) 309-8504
28 TELEFAX: (301) 309-8512
29
30 INFORMATION FOR SEQ ID NO: 531:
31 SEQUENCE CHARACTERISTICS:
32 LENGTH: 942 base pairs
33 TYPE: nucleic acid
34 STRANDEDNESS: double
35 TOPOLOGY: linear
36
37 US-08-781-986A-531

```

Query Match	4.7%; Score 47.4; DB 4; Length 942;
-------------	-------------------------------------

Matches 138; Conservative 0; Mismatches 151; Indels 0; Gaps 0

66 AGATATAGTACGTATACTTAAGCAACAGATGAGGTCATACAGCATCATATGG 125

Db 313 AGAAACGTA CTTACTACGCAATGAGCAATTAGATGAACAGAATAAAGTCACTTCTAT 3722

126 AATTGATATCGAGCTAAGAAATAATATTAGCTTACGCTTTGGCTGTAAGTGGATTCA 185

Db 373 AGATGATATAAGAGCTCTAACTAGTATCGTATATTGCATGTGGTGCAACTGGTCTGG 432

186 TAATGTATCTAACTTCCTGATGACTATTATAAGATAAGAGACTGCTGAGAGATTTA 245

Db 433 CATAGGAGAAATTATATGATGACATATTCATGAGAAGTATTTTAAAGAAGACGA 4922

246 TCAAGATATATGCTAATCTTCATCTGCACTATTAGGTGAATAATGATCAATTTC 305

Db 493 AACGCCATCAATGCTGATTTTGTATTTAGTTAATGGTGATTCAATGGAACTATGTT 552

QY 306 TAAAGATATGCGCAATGGTTTTTATAGAATGAACTGGATTTTGAAGCT 354

Db 553 AAAACAAGGACATACGCTTTATTTAAGAGAAGATTCTATTAAAGAT 601

Search completed: November 21, 2004, 08:44:25
Job time : 122 secs

	Query Match	100.0%;	Score 1008;	DB 6;	Length 1008;
	Best Local Similarity	100.0%;	Pred. No. 5.2e-171;		
	Matches 1008;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	ATGGTTATACAACTTAACACCTCGATGATAGAAAGTGGATATCCACCCGTTGAAAACCAATA	60		
Db	1	ATGGTTATACAAATTMACCCCTGATGAATGAAGTGGATATCCACCCGTTGAAAACCAATA	60		
QY	61	GCAGAGATATAGTACCTATCTACTAACTTTAAGCAAAACAGATGAGGGTCATACAGCATCA	120		
Db	61	GCAGAGATATAGTACCTATCTACTAACTTTAAGCAAAACAGATGAGGGTCATACAGCATCA	120		
QY	121	TATGGATTGAATATCGAGCTTAACAAAAATATAATTACTTACGCTTTGGCTGTAAAGCGT	180		
Db	121	TATGGAATGTAAATATCGAGCTTAACAAAAATATAATTACTTACGCTTTGGCTGTAAAGCGT	180		
QY	181	ATTGATATATGATCTTAAACCTCCGATGACATCTATTATAGAATAAAGACCTGTGAGAGA	240		
Db	181	ATTGATATATGATCTTAAACCTCCGATGACATCTATTATAGAATAAAGACCTGTGAGAGA	240		
QY	241	ATTTATCAGAGATATATGCTCTAATCTTTCACTGCACTATTAGGTGAAAATGCGATCA	300		

Db 241 ATTTATCAAGATATATGCTTAATCTTTTCATCTGCACTATATAGTGAAATGATGATCA 300
Qy 301 ATTTCTAAAGATATGCGCAAAATGTTTTTAAGAATGAACGTGATTTTGAAGCTCAATAT 360
Db 301 ATTTCTAAAGATATGCGCAAAATGTTTTTAAGAATGAACGTGATTTTGAAGCTCAATAT 360
Qy 361 CCTCAAAACATTTTGAATGTTCTCTGAGCTTGAATAAACCAATGAGTCTTATTCAGAT 420
Db 361 CCTCAAAACATTTTGAATGTTCTCTGAGCTTGAATAAACCAATGAGTCTTATTCAGAT 420
Qy 421 GACGATTAATATTTAGCACTATATTTTCTCTGTACAGAAATTCACATGAGGAAAT 480
Db 421 GACGATTAATATTTAGCACTATATTTTCTCTGTACAGAAATTCACATGAGGAAAT 480
Qy 481 CAACATCAAAATGCGCAGAGATTTTAAATTAATGATTTCTTAATTTACCTTATCTGCT 540
Db 481 CAACATCAAAATGCGCAGAGATTTTAAATTAATGATTTCTTAATTTACCTTATCTGCT 540
Qy 541 GTAACCTTCACTGGAGAGAGATTTTTCAAAACCTTTTACAAATGATTAAGGCTTAA 600
Db 541 GTAACCTTCACTGGAGAGAGATTTTTCAAAACCTTTTACAAATGATTAAGGCTTAA 600
Qy 601 TCATTAGAGAAATTAATATGAGAGAAAACCTTTCTAAACCTTTCTTGACACCGGAG 660
Db 601 TCATTAGAGAAATTAATATGAGAGAAAACCTTTCTAAACCTTTCTTGACACCGGAG 660
Qy 661 AGATTACCTGATGCGAGAAATAGTTATTTGCTGACCAAGAAACCGCTTAAATGAGA 720
Db 661 AGATTACCTGATGCGAGAAATAGTTATTTGCTGACCAAGAAACCGCTTAAATGAGA 720
Qy 721 GTGAGTTTTTAAAGAACTTAAATAACAATCTAGAAATGATTTTCTAATATGAGAGG 780
Db 721 GTGAGTTTTTAAAGAACTTAAATAACAATCTAGAAATGATTTTCTAATATGAGAGG 780
Qy 781 GCTGCAAAACAAAAGTATGATTCATTTTAAAGAGGTAACAAGGATGAGCTCCACAG 840
Db 781 GCTGCAAAACAAAAGTATGATTCATTTTAAAGAGGTAACAAGGATGAGCTCCACAG 840
Qy 841 ACAGCAGCGAAAAGTATGTTGTAACAGCCAGTGCAGTAACCTGAGAAAATTTGCCGATAT 900
Db 841 ACAGCAGCGAAAAGTATGTTGTAACAGCCAGTGCAGTAACCTGAGAAAATTTGCCGATAT 900
Qy 901 TTATATAGTGTAGGCTTAAGCCAAAAGACAGGGTAACCTTTTCTCAAAATGATCTGAC 960
Db 901 TTATATAGTGTAGGCTTAAGCCAAAAGACAGGGTAACCTTTTCTCAAAATGATCTGAC 960
Qy 961 AATACATGACGGTTCATAGTGTGGAACCTCATTTAATAAATATATGTA 1008
Db 961 AATACATGACGGTTCATAGTGTGGAACCTCATTTAATAAATATATGTA 1008

RESULT 2
AX029370 1008 bp DNA linear PAT 16-SEP-2000
LOCUS Sequence 2 from Patent WO903328.
AX029370
VERSION AX029370.1 GI:10190184
KEYWORDS
SOURCE .
ORGANISM Photorhabdus luminescens
Photorhabdus luminescens
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Photorhabdus.
REFERENCE
AUTHORS East, P.D.
TITLE Toxin genes from the bacteria xenorhabdus nematophilus and
JOURNAL photorhabdus luminescens
COMMUN SCIENT IND RES ORG (AU) Patent: WO 903328-A 2 28-JAN-1999;
FEATURES EAST PETER DAVID (AU)
source Location/Qualifiers
1..1008
/organism="Photorhabdus luminescens"
/mol_type="unassigned DNA"

ORIGIN /db_xref="taxon:29488"
Query Match 100.0%; Score 1008; DB 6; Length 1008;
Best Local Similarity 100.0%; Pred. No. 5.2e-171;
Matches 1008; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 ATGTTTATCAATTAACCTTGATGATGAGATATCCACCCGTTGAAAAGCAATA 60
Db 1 ATGTTTATCAATTAACCTTGATGATGAGATATCCACCCGTTGAAAAGCAATA 60
Qy 61 GAGAGATATATGATACGATATCTAACTTTAAGCAACAGATGAGGCTATACAGATCA 120
Db 61 GAGAGATATATGATACGATATCTAACTTTAAGCAACAGATGAGGCTATACAGATCA 120
Qy 121 TATGGAATTTGATATGAGCTAAGAAAATAATATTTAGCTTGGCTGTAAAGTGT 180
Db 121 TATGGAATTTGATATGAGCTAAGAAAATAATATTTAGCTTGGCTGTAAAGTGT 180
Qy 181 ATTCATTAATGATCTTAAACCTTCTGATGATCTATTATTAAGAATTAAGAGACTGAGAGA 240
Db 181 ATTCATTAATGATCTTAAACCTTCTGATGATCTATTATTAAGAATTAAGAGACTGAGAGA 240
Qy 241 ATTTATCAAGAAATATATGCTTAATCTTTCATCTGACCTAATTAAGGAGAAATGATCA 300
Db 241 ATTTATCAAGAAATATATGCTTAATCTTTCATCTGACCTAATTAAGGAGAAATGATCA 300
Qy 301 ATTTCTAAAGATATGCGCAAAATGTTTTTAAAGATGAACGTGATTTTGAAGCTCAATAT 360
Db 301 ATTTCTAAAGATATGCGCAAAATGTTTTTAAAGATGAACGTGATTTTGAAGCTCAATAT 360
Qy 361 CCTCAAAACATTTTGAATGTTCTCTGAGCTTGAATAAACCAATGAGTCTTATTCAGAT 420
Db 361 CCTCAAAACATTTTGAATGTTCTCTGAGCTTGAATAAACCAATGAGTCTTATTCAGAT 420
Qy 421 GACGATTAATATTAAGCACTATATTTTCTCTGTACAGAAATTCACATGAGGAGAAAT 480
Db 421 GACGATTAATATTAAGCACTATATTTTCTCTGTACAGAAATTCACATGAGGAGAAAT 480
Qy 481 CAACATCAAAATGCGCAGAAATTTTTTAAATTAATGATTTCTTAATTTACCTTATCTGCT 540
Db 481 CAACATCAAAATGCGCAGAAATTTTTTAAATTAATGATTTCTTATTTACCTTATCTGCT 540
Qy 541 GTAACCTTCACTGGAGAGAGATTTTTCAAAACCTTTTCAATGATTTAGAGCTTAA 600
Db 541 GTAACCTTCACTGGAGAGAGATTTTTCAAAACCTTTTCAATGATTTAGAGCTTAA 600
Qy 601 TCATTAGAGAAATTAATATGAGAGAAAACCTTCTAAACCTTTCTTGACACCGGAG 660
Db 601 TCATTAGAGAAATTAATATGAGAGAAAACCTTCTAAACCTTTCTTGACACCGGAG 660
Qy 661 AGATTACCTGATGCGAGAAATAGTTATTTGCTGACCAAGAAACCGCTTAAATGAGA 720
Db 661 AGATTACCTGATGCGAGAAATAGTTATTTGCTGACCAAGAAACCGCTTAAATGAGA 720
Qy 721 GTGAGTTTTTAAAGAACTTAAATAACAATCTAGAAATGATTTTCTAATATGAGAGG 780
Db 721 GTGAGTTTTTAAAGAACTTAAATAACAATCTAGAAATGATTTTCTAATATGAGAGG 780
Qy 781 GCTGCAAAACAAAAGTATGTTGTAACAGCCAGTGCAGTAACCTGAGAAAATTTGCCGATAT 840
Db 781 GCTGCAAAACAAAAGTATGTTGTAACAGCCAGTGCAGTAACCTGAGAAAATTTGCCGATAT 840
Qy 841 ACAGCAGCGAAAAGTATGTTGTAACAGCCAGTGCAGTAACCTGAGAAAATTTGCCGATAT 900
Db 841 ACAGCAGCGAAAAGTATGTTGTAACAGCCAGTGCAGTAACCTGAGAAAATTTGCCGATAT 900
Qy 901 TTATATAGTGTAGGCTTAAGCCAAAAGACAGGGTAACCTTTTCTCAAAATGATCTGAC 960
Db 901 TTATATAGTGTAGGCTTAAGCCAAAAGACAGGGTAACCTTTTCTCAAAATGATCTGAC 960
Qy 961 AATACATGACGGTTCATAGTGTGGAACCTCATTTAATAAATATATGTA 1008
Db 961 AATACATGACGGTTCATAGTGTGGAACCTCATTTAATAAATATATGTA 1008

Db 961 AATACATGACGGTTCATAGTGTGGAACTGATTAATAAATATATGA 1008

RESULT 3
LOCUS AX029374 1388 bp DNA linear PAT 16-SEP-2000
DEFINITION Sequence 6 from Patent WO9903328.
ACCESSION AX029374
VERSION AX029374.1 GI:10190186
KEYWORDS
SOURCE Photorhabdus luminescens
ORGANISM Photorhabdus luminescens
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Photorhabdus.

REFERENCE
AUTHORS East, P. D.
TITLE Toxin genes from the bacteria xenorhabdus nematophilus and photorhabdus luminescens
JOURNAL Patent: WO 9903328-A 6 28-JAN-1999;
COMMUN SCIENT IND RES ORG (AU) ; EAST PETER DAVID (AU)
FEATURES
source location/Qualifiers
1..1388
/organism="Photorhabdus luminescens"
/mol_type="unassigned DNA"
/db_xref="taxon:29488"

ORIGIN
Query Match 100.0%; Score 1008; DB 6; Length 1388;
Best Local Similarity 100.0%; Pred. No. 4,7e-171;
Matches 1008; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGTTATACATTAACCTGATGATAGTAATGATATCCACCGGTGAAGAAATGA 60
DB 172 ATGGTTATACATTAACCTGATGATAGTAATGATATCCACCGGTGAAGAAATGA 231
QY 61 GCAGAGATATATAGTACGTATACCTTAAGCAACAGATGAGGTCTATACAGATCA 120
DB 232 GCAGAGATATATAGTACGTATACCTTAAGCAACAGATGAGGTCTATACAGATCA 291
QY 121 TATGGAATGAAATGAGCTAAGAAATATATATAGCTTACGCTTGGCTGTAAGTGT 180
DB 292 TATGGAATGAAATGAGCTAAGAAATATATATAGCTTACGCTTGGCTGTAAGTGT 351
QY 181 ATTCAATATGATCTAACTTCCGATGATCTTTATTAAGAAATGAAGATGAGCTGAGAGA 240
DB 352 ATTCAATATGATCTAACTTCCGATGATCTTTATTAAGAAATGAAGATGAGCTGAGAGA 411
QY 241 ATTATCAAGATATATATGCTATCTTTCATCTGCACTATTAAGTGAATGATGATCA 300
DB 412 ATTATCAAGATATATGCTATCTTTCATCTGCACTATTAAGTGAATGATGATCA 471
QY 301 ATTCTTAAGATATGAGCAATGCTTTTATTAAGAAATGAAGATGAGCTGAGCTCAATAT 360
DB 472 ATTCTTAAGATATGAGCAATGCTTTTATTAAGAAATGAAGATGAGCTGAGCTCAATAT 531
QY 361 CCTGAAAACATTTGGAATGCTTCCGATGATGAGAAATGAAGATGAGCTTATCAAT 420
DB 532 CCTGAAAACATTTGGAATGCTTCCGATGATGAGAAATGAAGATGAGCTTATCAAT 591
QY 421 GACGATTAATATTTAGCACTATATTTTCTCTGTAAGGAATTCACATGAGAGAAAT 480
DB 592 GACGATTAATATTTAGCACTATATTTTCTCTGTAAGGAATTCACATGAGAGAAAT 651
QY 481 CAACATCAATATGCGCAAGATTTTAAATTAATGATTTCTTAATTAATCTTATCTGCT 540
DB 652 CAACATCAATATGCGCAAGATTTTAAATTAATGATTTCTTAATTAATCTTATCTGCT 711
QY 541 GTAACCTCACTGGAAGAGATTTTCAAAAACCTTTAAGAGATTAAGGCTAA 600
DB 712 GTAACCTCACTGGAAGAGATTTTCAAAAACCTTTAAGAGATTAAGGCTAA 771
QY 601 TCATTAGAGATTAATGAGAGAAAAAATCTTCTTTCGACCAACGCGAG 660

Db 772 TCATTAGAGATTAATGAGAGAAAAAATCTTCTTTCGACCAACGCGAG 831

QY 661 AGATTACCTGATGCGAATAGTATTTGGCTGAGCAAGAGCGCTTAATGAGA 720
DB 832 AGATTACCTGATGCGAATAGTATTTGGCTGAGCAAGAGCGCTTAATGAGA 891
QY 721 GTGAGTTTAAAGACTTAATAAATCAATCTAGATGATTTTCTTAATGAGAGG 780
DB 892 GTGAGTTTAAAGACTTAATAAATCAATCTAGATGATTTTCTTAATGAGAGG 951
QY 781 GCTGCAAAACAAAGTATAGTTCATTATTAAGAGGTACAAAGGGTAAAGCTCAAG 840
DB 952 GCTGCAAAACAAAGTATAGTTCATTATTAAGAGGTACAAAGGGTAAAGCTCAAG 1011
QY 841 ACAGAGCAAAAGATGATGATGATGACGCGATGAGTAACTGAGAAATTTGCCAATAT 900
DB 1012 ACAGAGCAAAAGATGATGATGATGACGCGATGAGTAACTGAGAAATTTGCCAATAT 1071
QY 901 TTATATAGTGTAGGCTTAAGCCAAAGACAGAGGTAACTTACTCAAAATGATAGTAC 960
DB 1072 TTATATAGTGTAGGCTTAAGCCAAAGACAGAGGTAACTTACTCAAAATGATAGTAC 1131
QY 961 AATACATGACGGTTCATAGTGTGGAACCTCAATTAATAAATATATGA 1008
DB 1132 AATACATGACGGTTCATAGTGTGGAACCTCAATTAATAAATATATGA 1179

RESULT 4
LOCUS BD080162 1007 bp DNA linear PAT 27-AUG-2002
DEFINITION Toxin genes from bacteria Xenorhabdus nematophilus and Photorhabdus luminescens.
ACCESSION BD080162.1 GI:22625765
VERSION JP 2001510022-A/2.
KEYWORDS Photorhabdus luminescens
SOURCE Photorhabdus luminescens
ORGANISM Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Photorhabdus.
1 (bases 1 to 1007)
REFERENCE
AUTHORS East, P. D.
TITLE Toxin genes from bacteria Xenorhabdus nematophilus and Photorhabdus luminescens
JOURNAL Patent: JP 2001510022-A 2 31-JUL-2001;
COMMENT COMMONWEALTH SCIENTIFIC AND INDUSTRIAL RESEARCH ORGANISATION
OS Photorhabdus luminescens
PN JP 2001510022-A/2
PD 31-JUL-2001
PF 17-JUL-1998 JP 2000502652
PR 17-JUL-1997 AU PO 8088
PI PETER DAVID EAST
PC C12N1/09,A01H5/00,A01N63/00,C07K14/24,C12N1/10,C12N1/19 PC
PC C12N1/21,C12N5/10
PC C12N7/00,C12N15/00,C12N5/00
CC Toxin genes from bacteria Xenorhabdus nematophilus and CC
CC Photorhabdus luminescens
CC luminescens
FH key
FT source
FEATURES
source location/Qualifiers
1..1007
/organism="Photobacterium luminescens".
/mol_type="genomic DNA"
/db_xref="taxon:29488"

ORIGIN
Query Match 99.9%; Score 1007; DB 6; Length 1007;
Best Local Similarity 100.0%; Pred. No. 7.8e-171;
Matches 1007; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 ATGGTTATACATTAACCTGATGATAGTAATGATATCCACCGGTGAAGAAATGA 60

Db 1 ATGTTTATACAACTTACACCTGATGATAGAGTGGATATCCACCCGTGAAAAACAAATA 60
Qy 61 GCAGAGATATATAGTACGTATACCTTAAAGCAAAACAGATGAGGGTCATACAGCATCA 120
Db 61 GCAGAGATATATAGTACGTATACCTTAAAGCAAAACAGATGAGGGTCATACAGCATCA 120
Qy 121 TATGGAATGATATCGAGCTTAAGAAATATATATAGCTTACGCTTGGCTGTAAGTGT 180
Db 121 TATGGAATGATATCGAGCTTAAGAAATATATATAGCTTACGCTTGGCTGTAAGTGT 180
Qy 181 ATTCATATATGATCTTAACTTCCTGATGACCTATTTATAGAAATTAAGAGCTCTGAGAGA 240
Db 181 ATTCATATATGATCTTAACTTCCTGATGACCTATTTATAGAAATTAAGAGCTCTGAGAGA 240
Qy 241 ATTTATCAAGATATATATGCTATATCTTTCATCTGCACTATTAGTGAATGATGATCA 300
Db 241 ATTTATCAAGATATATATGCTATATCTTTCATCTGCACTATTAGTGAATGATGATCA 300
Qy 301 ATTTCTAAAGATATGCGCAATGCTTTTATTAAGAAATGAATGAACTGGATTTGAAAGTCAATAT 360
Db 301 ATTTCTAAAGATATGCGCAATGCTTTTATTAAGAAATGAATGAACTGGATTTGAAAGTCAATAT 360
Qy 361 CCTCAAAACATTTGGAATGCTCTGAGCTTGAATTAACCAATGAGTGTCTATTACAGT 420
Db 361 CCTCAAAACATTTGGAATGCTCTGAGCTTGAATTAACCAATGAGTGTCTATTACAGT 420
Qy 421 GACGATTAATATATATAGCACTATATTTTCTCTGTACAGAAATTCACCTGAGAGAAAT 480
Db 421 GACGATTAATATATATAGCACTATATTTTCTCTGTACAGAAATTCACCTGAGAGAAAT 480
Qy 481 CAACAATCAATATGCGCAAGATTTTAAATTAATGATTTCTTATTTACCTTATCTGCT 540
Db 481 CAACAATCAATATGCGCAAGATTTTAAATTAATGATTTCTTATTTACCTTATCTGCT 540
Qy 541 GAAACTTCACTGAGGAAGAGATTTTCAAAAACTTTTACAATGAGATTAAGGCTAA 600
Db 541 GAAACTTCACTGAGGAAGAGATTTTCAAAAACTTTTACAATGAGATTAAGGCTAA 600
Qy 601 TCATTAAGAAATTAATATGAGAAAAAACTTTCTTAACTTTCTTTCGACACCGCAG 660
Db 601 TCATTAAGAAATTAATATGAGAAAAAACTTTCTTAACTTTCTTTCGACACCGCAG 660
Qy 661 AGATTACCTGATGCGAATAGGTTATTTGGCTGACCAACAGAGCGCTTAAATGAGAGA 720
Db 661 AGATTACCTGATGCGAATAGGTTATTTGGCTGACCAACAGAGCGCTTAAATGAGAGA 720
Qy 721 GGGATTTTAAAGACTTAATAAACAATCTAGGAATGAGATTTCTTAAATGAGAGGG 780
Db 721 GGGATTTTAAAGACTTAATAAACAATCTAGGAATGAGATTTCTTAAATGAGAGGG 780
Qy 781 GCTGCAAAACAAAGATATAGTTCATTTATTAAGAGGTACAAAGGGTAACGCTCCACAG 840
Db 781 GCTGCAAAACAAAGATATAGTTCATTTATTAAGAGGTACAAAGGGTAACGCTCCACAG 840
Qy 841 ACAGACGCGAAAAAGTATGCTACAGCGAGTGGCACTGAGGAAAAATTCGCCGAATAT 900
Db 841 ACAGACGCGAAAAAGTATGCTACAGCGAGTGGCACTGAGGAAAAATTCGCCGAATAT 900
Qy 901 TTATATAGTGTAGGCTAAGCAAAAGACAGGTTAACTTTAACTCAAAATATATCTGAC 960
Db 901 TTATATAGTGTAGGCTAAGCAAAAGACAGGTTAACTTTAACTCAAAATATATCTGAC 960
Qy 961 AATACATGACGCTTATAGTGTGGAATCTATTAATAAATAATATATG 1007
Db 961 AATACATGACGCTTATAGTGTGGAATCTATTAATAAATAATATATG 1007

RESULT 5
LOCUS BX571866 349652 bp DNA linear BCT 26-SEP-2003
DEFINITION Photorhabdus luminescens subsp. laumondii TT01 complete genome;
ACCESSION BX571866 BX470251

VERSION BX571866.1 GI:36785405
KEYWORDS complete genome.
SOURCE Photorhabdus luminescens subsp. laumondii TT01
ORGANISM Photorhabdus luminescens subsp. laumondii TT01
REFERENCE 1 Duchaud,E., Rusniok,C., Frangeul,L., Buchrieser,C., Taourit,S.,
Bocs,S., Boursaux-Bude,C., Chanderler,M., Daasa,E., Deroose,R.,
Derzelle,S., Freysinet,G., Gaudreau,S., Givaudan,A., Glaeser,P.,
Medigue,C., Lanois,A., Powell,K., Signier,P., Wingate,V.,
Zouine,M., Boemare,N., Danchin,A. and Kunst,F.
Complete genome sequence of the entomopathogenic bacterium
Photorhabdus luminescens
Nat. Biotechnol. 11 (1) (2003) In press
2 Duchaud,E., Frangeul,L., Rusniok,C. and Kunst,F.
Direct Submission
Submitted (23-APR-2003) L. Frangeul, Institut Pasteur, Genopole, 25
rue du Docteur Roux, 75724 Paris Cedex 15, FRANCE. E-mail:
lfrangeul@pasteur.fr, fkunst@pasteur.fr
location/Qualifiers
1. 349652
/organism="Photorhabdus luminescens subsp. laumondii TT01"
/mol_type="genomic DNA"
/strain="TT01"
/db_xref="taxon:243265"
572. .2086
/locus_tag="plu2059"
572. .577
/locus_tag="plu2059"
584. .2086
/locus_tag="plu2059"
/function="Unknown"
/note="unnamed protein product; Highly similar to unknown
protein of Photorhabdus"
/codon_start=1
/transl_table=11
/protein_id="CAE14352.1"
/db_xref="GI:36785406"
/translation="MSILTDIVSGITVLEACPVIITVTSAGDPTHKVETETDIWYSH
PKATRFDPKTRILADVENNTCTAKFVTVTASDADGFTITFSDMKATTPDADLGIAPHS
IDKDFSHAEPEPILIGDSLTKDLSPVADDERPSEFNSGLINITALSKTNEPLPYFO
IPVINDLVRIFGYDGENITBELIPEIKIINNKKYIINTDQKGVALKVPRKKEKIGN
YVPMFIELPGLKNAASNIILFIEDTAPALPADIIBELDNRLTPSVGNSGTGPHVS
IPFYKAKPRDFLIVREDNDELGASTVITDLYOLANPFLIVPAGVPIGDNELC
YVVDVVCNVMSSALYENLIGSVNPEPEKRTILAVPRIAHHKQOLEFQSGRLNE
NIIIGGLNAHIAVGKNGIDVDVSVINRVIYNGVINDPSSRLVSLDPTHTYADBEVAA
GETIIQLQAPFPITKVGSTCDGTPQSVVIYTVSDTKESIMSGHINTAFVG"
2442. .3971
/locus_tag="plu2060"
2442. .3971
/locus_tag="plu2060"
/function="Unknown"
/note="unnamed protein product; Highly similar to unknown
protein of Photorhabdus"
/codon_start=1
/transl_table=11
/protein_id="CAE14353.1"
/db_xref="GI:36785407"
/translation="WFENMRFPNMSILTDIVSGITVLEACPVIITVTSAGDSEHKVT
ETTDIVSDQDVSEFDETPRI.SRDVKNNICRAKFPVVAHSAVANDFTIIFSDKTTDP
AKLEIAFPIIDKNGFSAHAFPMIIGSLVNDLPVADDEPSEFNSGLIYINILANK
TDDPLPYFOILPLVTDKRVRIFGYBEGQTEIFPFIOLANSKYIINTNSGFGALRA
FPGSGEGNVVMNFTLQGRSSSNILFITDIAPIIAMPDIKEIDGRILIPSVG
NSSTGFHVSIFPYEGAKKDRIFLPIAIRDNDELIGSTSVITDINDLSDPISAPYASI
PNIGDNEIYVADTMGNVAISSILYFTLTVSVNPQPEIRITLAAAKIFHKDSEL
KORGKINLVKILGGGLDARIAVGKNGIDVDVSVINRVIYNGVINDPSSSKITSLSEPH
TVTAEEVYAGRAIILLGQAPFTGVSDYDGGGLVYVTVSVNIKNSKIFGRIDTVPG
O"
4296. .5909
/locus_tag="plu2061"
4296. .4302
gene
RBS

CDS

1/locus_tag="plu2061"
4311..5909
/locus_tag="plu2061"
/function="unknown"
/note="unnamed protein product; Highly similar to unknown protein of Photorhabdus"
/codon_start=1
/transl_table=1
/protein_id="CAE14354.1"
/db_xref="GI:36785408"
/translation="MMKICNFDKRLDVFYLLMFNPNRLNMTLADIVSGITILEGC
PVIIVYSGEYVHKVFEATDIYVSDOODVSFDEIPIRSDVKNIYTAKEVVASD
ASDFIMSSMKAQDPDPAKLEIPEHSGIKGNDPSYEGPTPIIGDSMTKDSPEQDNR
PSLFRDLLINLYALNKNDPLPFQPIPIYDQVRIFGNRQIIREILPIFSEKON
COYVINTDSKGLALIVEPKSGEYVNMILTKPLYLKSSASSNLTITDIPETLPA
DIALNKNKSPVSGNSNTDFVHIIPEYEGVRRDVFALIIRDDEDTQGAASVITD
LNQDMSTISVYANIPTNGNNEIYYVADAGNVAASARFYLTKPVLNVPETIR
ILAVRISIKIDQVFEROSSKINLASIIGGLDAHIAVGEKNGEINNTVNWIIYNG
IINAEPEFKISIALPHYTLADVMAGEAIIPLGSAITFNNSYRQDGPVNIYTSVG
MECSSTHSHKIVGRIDITVPPGR"
Join(5974..5979,5993..6679)
1/locus_tag="plu2062"
5974..5993
1/locus_tag="plu2062"
5993..6679
/locus_tag="plu2062"
/function="unknown"
/note="unnamed protein product; Some similarities with unknown protein of Photorhabdus"
/codon_start=1
/transl_table=1
/protein_id="CAE14355.1"
/db_xref="GI:36785409"
/translation="MNNGISVSVIITENAAKMSDLTDYDNPINLPONIVITLGHGPKHFEYAT
TSIGAFYDDNINLDOIILITLDEGKGHFYVRSPEHKNIKSEPSAFVIAPOEII
NKMVSPFLFGYVROSDEALITETAYNTYTGAPADQKPCSTIYLFIDREHNDIDNOIRI
RVNNNALIDGYNKMDIPLKEDSAYTVNIVSNYGVKNWLTLPDSDSQKYNFVLS
FRPTMGSEI"
6664..6669
1/locus_tag="plu2063"
6664..6669
/locus_tag="plu2063"
6676..8001
1/locus_tag="plu2063"
6676..8001
/locus_tag="plu2063"
/function="unknown"
/note="unnamed protein product; Some similarities with unknown protein of Photorhabdus"
/codon_start=1
/transl_table=1
/protein_id="CAE14356.1"
/db_xref="GI:36785410"
/translation="WITNNDNTPISLSSEPELPQAEHGLDIVSQONODILIKYVTHE
WLRIDMIVVNGGIFSPRPYIIONINLDFVTVIIPFSEIPDGCNVTYTSADPFENEN
KSPVAVIKMISGKIFSLPSIYDPSISGVHISLVQHSGLNVAHSYQOMQKGSYVLF
HNGGNFDFPIARAAADQPIVSDDEPKQGYSLVVPANVLCLQEGKGGYEVF
GNSSAPGGSVILSLTDMSSAMISISEGAPWVSNRPINPKNLAVVGPGRKLSASV
EGSGIVBEGGAKQYSPNSRGLATPHLOSEVAGSFRCTVSAHEDSITTPGVSQAVFSW
RQGNISIVATRTSEVPADGVSCGYIVVRIVYSGRLRLVTDGSAIISGIDGNER
KQVVISKQCGTAFHITVTKAKNNVHISLPDTDKSIQPSINFGHPFV"
8105..9084
1/locus_tag="plu2064"
8105..9110
1/locus_tag="plu2064"
9116..9084
/locus_tag="plu2064"
/function="unknown"
/note="unnamed protein product; weakly similar to putative invasins and adhesins"
/codon_start=1
/transl_table=1
/protein_id="CAE14357.1"

gene

RBS

CDS

gene

RBS

CDS

```

/db_xref="GI:36785411"
/translacion="MNDEQAEKAKTLNMLSSSEKTADKPIDIGVAIODEYLHLGV
NIITFLDNQSMVFKEKNDRKIIEKGTVLGAIXSVELISDSTPCOGNVYACLEDPEIE
APILQVEFTVAPQSVAINTLNDLEKMAATKAGDENKVAVATYDRLGNGIKNOETFYA
DSGATVVSPSGMTDKDGKITIASIOSRAGEVILMAQAIAEKALMCFEYNYLEKTI I
EPFAYLVFSFGIILAGLYKNGVEPNNTIVSMSYSTLSODGHMTDDKNGPSFR
VSSSRGVSRYSHNHVPMSSEFSVYISCIDGRVYSQAWIELPLGLY"
9552. .9905
/locus_tag="piu2065"
9552. .9557
/locus_tag="piu2065"
9564. .9839
/locus_tag="piu2065"
/function="unknown"
/note="unnamed protein product; Similar to probable
membrane protein YchH of Escherichia coli"
/codon_start=1
/transl_table=1
/protein_id="CAE14358.1"
/db_xref="GI:36785412"
/translacion="MKROSARITVGNNVLMGVGFITYSSLIYVTLPHIFGPGNPELSE
AALGVPIGALVWLAVGAGVGGREVDARYMWLRNDERYRRGSHRP"
9881. .9905
/locus_tag="piu2065"
complement(9935. .10896)
/gene="prsa"
/locus_tag="piu2066"
complement(9935. .10882)
/gene="prsa"
/locus_tag="piu2066"
/codon_start=1
/transl_table=1
/product="Ribose-phosphate pyrophosphokinase
(phosphoribosyl pyrophosphate synthetase)"
/protein_id="CAE14359.1"
/db_xref="GI:36785413"
/translacion="MPDMKLPRAGNAIPELAQRVANRLYTNLGDAAVGRSDDEVSYOV
NEINRGDTIFTIOSTCAPTNDNLMEVLVWMDLKRASAGRITAVTPYFYSRADORRVR
SARPVTARKVADPLSVSGVDNRVLTVDLHAQIQGFPPVDNVGSPILLBMDLOXE
LENIVIVSPDIGVVARAIKALINDTDMAIIDKRPRANVSQWHIIIDVAGRDCVL
VDNDIDRGCTCKAABALKERGAKEPVFAVATHPIFGNAVVDVIKNSVIDEVVCDTI P
LSARIKLNKTRTITLSGMTAEAIRRISNESISMFEH"
complement(10891. .10896)
/gene="prsa"
/locus_tag="piu2066"
complement(10946. .11821)
/gene="ispB"
/locus_tag="piu2067"
complement(10946. .11821)
/gene="ispB"

CDS
Query Match      98.4%; Score 992; DB 1; Length 349652;
Best Local Similarity 99.0%; Pred No. Be-169;
Matches 998; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

gene
RBS
DB
1 ATGGTTATACATTAAACACTGTGATGATGAAGTGATATCCACCCTGAAAAACAATA 60
306840 ATGGTTATACATTAAACACTGTGATGATGAAGTGATATCCACCCTGAAAAACAATA 306899
Db
61 GCAGAGATATATGTCGTATACTTAACCTTTAAGCAAACAGATGAGGTCATACAGATCA 120
306900 GCAGAGATATATGTCGTATACTTAACCTTTAAGCAAACAGATGAGGTCATACAGATCA 306959
Oy
121 TATGGAATTGATATCGAGCTAAGAANAATATATAGCTTAGCTTTGCCTGAATGTGT 180
Db
306960 TATGGAATTGATATCGAGCTAAGAANAATATATAGCTTAGCTTTGCCTGAATGTGT 307019
Oy
181 ATTCAATAATGATCTAAACCTTCCTGATGACTATTTATTAAGATTAAGAAGACTGCTGAGA 240
Db
307020 ATTCAATAATGATCTAAACCTTCCTGATGACTATTTATTAAGATTAAGAAGACTGCTGAGA 307079
Oy
241 ATTTATCAAGATATATAGCTTAATCTTTCAATCTGCACTATTAGCGAANAATGCTGATCA 300

```

Db	307080	ATTATTCAAGATTATATGTCATCTTATCTGCACTATTAGTGAATAATGTTGATCA	307133
Qy	301	ATTCTAAGATATGGCAAAATGGTTTTTATPAAGATGAACGTGATTTTGAGGTCATAAT	360
Db	307140	ATTTCTAAGATATGGCAAAATGGTTTTTATPAAGATGAACGTGATTTTGAGGTCATAAT	307199
Qy	361	CCTCAAAACATTTGGAAATGTCCTGAGCTTGAAAAACCATGAGTCTTATTCAGAT	420
Db	307200	CCTCAAAACATTTGGAAATGTCCTGAGCTTGAAAAACCATGAGTCTTATTCAGAT	307259
Qy	421	GACGATTAATTTATGCACTATATTTTTCTCTGTACAGAAATTCACCTGAGAGAAAT	480
Db	307260	GACGATTAATTTATGCACTATATTTTTCTCTGTACAGAAATTCACCTGAGAGAGAAAT	307319
Qy	481	CAACATCAAAATGCCGCAAGATTTTTTAAATTAATTAATGATTTCTTATTTACCTTATGCT	540
Db	307320	CAACATCAAAATGCCGCAAGATTTTTTAAATTAATTAATGATTTCTTATTTACCTTATGCT	307379
Qy	541	GTAACCTTCATGGGAAGAGATTTTTCAAAAACCTTTTACAATGATTTAGAGGCTAAA	600
Db	307380	GTAACCTTCATGGGAAGAGATTTTTCAAAAACCTTTTACAATGATTTAGAGGCTAAA	307439
Qy	601	TCATTAGAAATTAATTATGAGAGAAAAAAACCTTCAAAACCTTCTTTGACACACGCGAG	660
Db	307440	TCATTAGAAATTAATTATGAGAGAAAAAAACCTTCTTAAACCTTCTTTGACACACGCGAG	307499
Qy	661	AGATTACCTGATGGCAGAAATAGGTTATTTGGCTGACCAACAGAACGCTTAATAGAGAT	720
Db	307500	AGATTACCTGATGGCAGAAATAGGTTATTTGGCTGACCAACAGAACGCTTAATAGAGAT	307559
Qy	721	GTGAGTTTTTAAAGAACTTAAAAATTAACAATCTAGAAATGGAATTTCTTAATATGAGAGG	780
Db	307560	GTGAGTTTTTAAAGAACTTAAAAATTAACAATCTAGAAATGGAATTTCTTAATATGAGAGG	307619
Qy	781	GCTGCAAAACAAAGATATGTTCACTTATPAAGAGGTCAAAAAGGTTAACGCTCACAG	840
Db	307620	GCTGCAAAACAAAGATATGTTCACTTATPAAGAGGTCAAAAAGGTTAACGCTCACAG	307679
Qy	841	ACAGACGCAAAAAGTATTTGTTACAGCCAGTGGCAGTACCTGGAATAATTTGCCGATTAAT	900
Db	307680	ACAGACGCAAAAAGTATTTGTTACAGCAACGCGAGTACCTGGAATAATTTGCCGATTAAT	307739
Qy	901	TTATATAGTGTAGGCTTAAGCCAAAAAGACAGGTAACCTTATCTCAAAATGATCTGAC	960
Db	307740	TTATATAGTGTAGGCTTAAGCCAAAAAGACAGGTAACCTTATCTCAAAATGATCTGAC	307799
Qy	961	AATACAATGACGTTCACTAGTGTGGAACTCATTAATAAAATTTATGA	1008
Db	307800	AATACAATGACGTTCACTAGTGTGGAACTCATTAATAAAATTTATGA	307847
RESULT 6			
LOCUS	AX770906/c		
DEFINITION	Sequence 37 from Patent WO02094867.	DNA	linear
ACCESSION	AX770906		
VERSION	AX770906.1		
KEYWORDS	GI:32438070		
SOURCE			
ORGANISM	Photorhabdus luminescens		
REFERENCE	Photorhabdus luminescens		
AUTHORS	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Photorhabdus.		
TITLE	1 Duchaud, E., Taourit, S., Glaeser, P., Frangeul, L., Kunst, F., Danchin, A. and Buchrieser, C.		
JOURNAL	Sequence of the Photorhabdus luminescens strain T701 genome and		
FEATURES	Patent: WO 02094867-A 37 28-NOV-2002;		
source	INSTITUT PASTEUR (FR) ; CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE (CNRS) (FR)		
	Location/Qualifiers		
	1. .349980		
	/organism="Photorhabdus luminescens"		

Query Match	Best Local Similarity	98.4%	Score 992	DB 6	Length 349980
Matches 998	Conservative 0	Mismatches 10	Indels 0	Gaps 0	
1	ATGGTTATACATTAACA	CGTGAATATGAAGTGAATGATCCACCCGTTGAAAAAGCAATA	60		
166220	ATGGTTATACATTAACA	CGTGAATATGAAGTGAATGATCCACCCGTTGAAAAAGCAATA	161616		
61	GCAGAGATATATAGTAC	GTATACCTTAACCTTTAAGCAACAGATGAGGGTATACAGCATCA	120		
166160	GCAGAGATATATAGTAC	GTATACCTTAACCTTTAAGCAACAGATGAGGGTATACAGCATCA	166101		
121	TATGGAATTTGAATATG	AGCTAAGAAATATATTTAGCTTACGCTTGGCTGTAACTGT	180		
166100	TATGGAATTTGAATATG	AGCTAAGAAATATATTTAGCTTACGCTTGGCTGTAACTGT	166041		
181	ATTCTAATATATCTAA	CTCCGATGATCTATTTAATGAATTAAGAGAGCTGTGAGAG	240		
166040	ATTCTAATATATCTAA	CTCCGATGATCTATTTAATGAATTAAGAGAGCTGTGAGAG	165981		
241	ATTATTCAGAAATATAT	ATGCTATATCTTATCTGACCTATTAGTGGAATATGATGATCA	300		
165980	ATTATTCAGAAATATAT	ATGCTATATCTTATCTGACCTATTAGTGGAATATGATGATCA	165921		
301	ATTCTAAGATATGCGAA	ATGCTTTTATTAAGAAATGAACCTGATTTTGAAGGTCATAT	360		
165920	ATTCTAAGATATGCGAA	ATGCTTTTATTAAGAAATGAACCTGATTTTGAAGGTCATAT	165861		
361	CCTCAAAATTTTGAAT	GTCTCCATGCTGAAATTAACCTTGAAGCTTATTCAGAT	420		
165860	CCTCAAAATTTTGAAT	GTCTCCATGCTGAAATTAACCTTGAAGCTTATTCAGAT	165801		
421	GACGATTAATTTTGA	CACTATATTTTCTGTGACAGGAATTCACCTGAGAGAAAT	480		
165800	GACGATTAATTTTGA	CACTATATTTTCTGTGACAGGAATTCACCTGAGAGAAAT	165741		
481	CAACATCAAAATGCG	CAAGATTTTATTAATTAATGATTTCTTATTTA	540		
165740	CAACATCAAAATGCG	CAAGATTTTATTAATTAATGATTTCTTATTTA	165681		
541	GTAACCTCACTGGGA	AGGATTTTTTCAAAAACTTTTACATGAGATTAGAGGCTTAA	600		
165680	GTAACCTCACTGGGA	AGGATTTTTTCAAAAACTTTTACATGAGATTAGAGGCTTAA	165621		
601	TCATTAAGAAATTA	TATGAGAGAAAACTTTCTTAACTTTTCTTGA	660		
165620	TCATTAAGAAATTA	TATGAGAGAAAACTTTCTTAACTTTTCTTGA	165561		
661	AGATTACCTGATG	CGAGATAGGTTATTTGCTGACCAACAGAGCGCTTAATGAGA	720		
165560	AGATTACCTGATG	CGAGATAGGTTATTTGCTGACCAACAGAGCGCTTAATGAGA	165501		
721	GTGAGTTTAAAGAA	CTTTAAAAATTAACAATCTAGGAATGATTTTCTTAATATGAAAGG	780		
165500	GTGAGTTTAAAGAA	CTTTAAAAATTAACAATCTAGGAATGATTTTCTTAATATGAAAGG	165441		
781	GCTGAAAAAACA	AACTATATGTTCAATTTAATAAGAGGTCAAAAGGTTAACGCTCCACAG	840		
165440	GCTGAAAAAACA	AACTATATGTTCAATTTAATAAGAGGTCAAAAGGTTAACGCTCCACAG	165381		
841	ACAGACGGAAGAA	AGTATGTTGTCACAGCCAGTGGCAGATTAACCTGGAATAATTTGCCAGTAAT	900		
165380	ACAGACGGAAGAA	AGTATGTTGTCACAGCCAGTGGCAGATTAACCTGGAATAATTTGCCAGTAAT	165321		
901	TTATATAGTGTAGG	CTTAAGCCAAAGACAGGGTAACTTTTACTCAAAATGATCTGAC	960		

[illegible]

[illegible]

AUTHORS	East, P. D.
TITLE	Toxin genes from bacteria Xenorhabdus nematophilus and Photorhabdus luminescens
JOURNAL	Patent: JP 2001510022-A 1 31-JUL-2001; COMMONWEALTH SCIENTIFIC AND INDUSTRIAL RESEARCH ORGANISATION
COMMENT	OS Xenorhabdus nematophilus PN JP 2001510022-A/1 PD 31-JUL-2001 PF 17-JUL-1998 JP 2000502652 PR 17-JUL-1997 AU PO 8088 PI PETER DAVID EAST PC C12N1/21, C12N5/10, A01H5/00, A01N63/00, C07K14/24, C12N1/10, C12N1/19 PC , C12N1/00, C12N15/00, C12N5/00 CC Toxin genes from bacteria Xenorhabdus nematophilus and CC Photorhabdus luminescens
FEATURES	CC key source location/Qualifiers FT 1..1107 /organism='Xenorhabdus nematophilus'. location/Qualifiers 1..1107 /organism='Xenorhabdus nematophila' /mol_type='genomic DNA' /db_xref='taxon:628'
ORIGIN	
Query Match	65.9%; Score 663.8; DB 6; Length 1107;
Best Local Similarity	79.4%; Pred. No. 2,6e-109;
Matches	835; Conservative 0; Mismatches 162; Indels 54; Gaps 2
OY	12 ATTAAACCTGATGATAGAGAAGTGATATCAACCGTTGAAGAACAAATPAGCAGAGATAT 71
DB	57 AGTAACGGCTGATGATAAAGGTGAATATCAACCGTGTAAAAACAATPAGCGGAGATAT 116
OY	72 AGTAGATATCTAACTTTAAGCAAACAGTAGAGGTGATACAGATCATATNGAATTGA 131
DB	117 AATAGGTATCTAGAAATTCAGCAACAAATGAATACAGATTTGTATGAAATTC 176
OY	132 ATATGAGCTAAGAAATATATATGCTTACCGTTGGCTGPAATGGTATTCATATATGT 191
DB	177 ATATGAGCTAAGAAATATATATGCTTACCGTTGGCTGPAATGGTATTCATATATGT 236
OY	192 ATCTAACTTCTGATGACTATATTAAGATTAAGAGA CTGCTGAGAAATTTATCAAGA 251
DB	237 CTCGAACCTTCCAAGAACTATATTAATAATPAGATTAACAGTAGAATTTATCAAGA 296
OY	252 ATAATGCTTAATCTTTGATGTCACATTTAGGTGAAAATGCTGATCAATTTCTTAAGA 311
DB	297 ATACATGCTTAATCTTTATCTGCACCTATGGGTGAGATGCTGATCAAAATTTCTTAAGA 356
OY	312 TATGGCAATGGTTTTTAAGAAATGAAGACTGSAATTTGAAGGTCAATATCCCAAAACT 371
DB	357 TATGGCAATGATTTTATCCAGAAAGAACTGAGTTTGAGGTCAACGTTTAAAAATAC 416
OY	372 TTGAATGTTCTGTGCTTGAATAATTAACATTTGATGTTTACATGACATTAAT 431
DB	417 CTGGATATTTCTGATCTTTGAGAAATTAATCTATTTGAAGATTAATTTGAGTAGAATTAAT 476
OY	432 ATTAGCACTATATTTTTCTCTGACAGAAATTCACCTGAGAGAAATCAACAATCAAA 491
DB	477 ATTAGCACTATATTTCTTTGCTTCAACAAGAACTTCCAAATGAGGCAAAATCAACAATCAAA 536
OY	492 TGCCGCCAAGATTTTAAATTAATTAATTTCTATTTACCTATATGCTGTAACCTTCT 551
DB	537 TGCCGCCAATTTTAAAGTAATTAATTTTACTATATCTTATGCTGTAACATCACT 596
OY	552 GGGAAGAGAGATTTTCAAAAACTTTTCAATCAATGATAGAGGCTAATCAATTAAGAGA 611
DB	597 GGGAAGAGAGATTTTCAAAAACTTTTCAATCAATGATAGAGGCTAATCAATTAATCAATTAAGAGA 656
OY	612 TTATATGAGAGAAAAAACTTTCTAAACCTTTCTTTGACCACCGCAGAGATTACTTGA 671

Db 657 TTATATTGAGAGAAAAAACTTTCTAACTTTCTTTCGACACCGCAGAACTTACCTGA 716
Qy 672 TGGGAGAAATAGATTATTTGGCTGACCAAGAGCGGCTAAA----- 714
Db 717 TGGGAGAAACAGGCTACTTGGCCGGTCCAAAGAGCGCTTAATTTGCCAAGACGCTTTC 776
Qy 715 -----TGGAGTGAAGTTTAAAGAACTTAA 740
Db 777 TACAGCAACAAAGCTTACAGCAGCTTCACTTAATTTGAGAGTTAGTTTGCAGAAACCTTAC 836
Qy 741 AAATAACAAATCTAGAAATGATTTCTTAATAAGAGGGGCTGCAAAACAAAGTATAG 800
Db 837 AGATPACCCATCCAGAAATACATTTATGAAATGATGATGCTGCAAAACGAAATATAG 896
Qy 801 TTCAATTATTAAGAGGTACAAAAGGTTAAGCTTCAACAGACAGAGGAGAAAGATTGG 860
Db 897 TTCAATTATTAAGAGGTACAAAAGGTTAAGTATCCAGTGCAGACAGACAGAGATTGG 956
Qy 861 TACAGCCAGTGGCAGTAACTGGAATAATGCGCAATA--ATTATATAGTGAAGGCT 917
Db 957 TACAAAAGCGGCACTTACCTTGCAGAACTGCAAGTATGAGATTTATATAGTAACT 1016
Qy 918 AAGCCAAAAGACAGGGTAACTTCTCAAAATGATATGACAAATACATGACGTTCA 977
Db 1017 AAGCCAAAAGACAGGGTAACTTCTCAATTAATATATATGACCAATATATGAGATCCA 1076
Qy 978 TAGTGTGGAATCTTATTAATAATATATGA 1008
Db 1077 AAGTGTGGAATCTTATCCAAATATATATAA 1107

RESULT 10

AR407889 LOCUS AR407889 1205 bp DNA linear PAT 18-DEC-2003
DEFINITION Sequence 5 from patent US 6630619.
ACCESSION AR407889
VERSION AR407889.1 GI:40157857
KEYWORDS

SOURCE

ORGANISM

Unknown.
Unclassified.

REFERENCE 1 (bases 1 to 1205)

AUTHORS

East, P.D.
Toxin genes from the bacteria *Xenorhabdus nematophilus* and

TITLE

Photorhabdus luminescens
Patent: US 6630619-A 5 07-OCT-2003;

JOURNAL

Location/Qualifiers

FEATURES

1..1205
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN

Query Match 65.9%; Score 663.8; DB 6; Length 1205;
Best Local Similarity 79.4%; Pred. No. 2.5e-109;
Matches 835; Conservative 0; Mismatches 162; Indels 54; Gaps 2;

Qy 12 ATTAACACCTGATGATGAAGTATTCACCCGTTGAAGCAATAGCAGAGATAT 71
Db 73 AGTAACGCGCTGATGATGAAGTATTCACCCGTTGAAGCAATAGCAGAGATAT 132
Qy 72 AGTAGTATCTAACTTTAAGCAACAGATGAGGTCTATACAGATCATATGAAATGA 131
Db 133 AATAGGTGATCTAGAAATTCAGCAACAAATGAAAGTCTACAGAGATGATGAAATGC 192
Qy 132 ATATGAGCTAAGAAATATATATAGCTTACGCTTGGCTGTAGTGTATTCATATGT 191
Db 193 ATATGAGCTAAGAAATATATATAGCATATGCTTTAGCGGTAAGTGTATTCATATGT 252
Qy 132 ATCTAACTTCTGATGATCTATTAAGAAATTAAGAGACTGCTGACAGAAATTTACAGA 251
Db 253 CTCTCAACTTCCAGAGACTATTAATAAATTAAGATTAACAGAGGATGAAATTTACAGA 312
Qy 252 ATATATGCTAATCTTTCATCTGCACTATATAGTGAAATGTGATCAATTTCTAAAGA 311

Db 313 ATACATGCTAATCTTTATCTGACATTTGGGAGAAATGATCAATTTCTAAAGA 372
Qy 312 TATGGCAAAATGTTTATATAGAAATGAACGGAATTTTGAAGTCAATCTCAAAACAT 371
Db 373 TATGGCAAAATGTTTATCCAGAAACGAGCTGAGAGTCAAGCTTTAAATAATAC 432
Qy 372 TTGGAATGTTCTGAGCTGAAATTAACATGAGTCTATTCAGATGACGATTAATTT 431
Db 433 CTGGATATTTCTGATCTTGAATTAACATTTATGGAAGATTTATCAGATGAAGTAAAT 492
Qy 432 ATTAGCACTAATATTTTCTCTGACAGAAATTCATGAGAGAAATCAACATCAAA 491
Db 493 ATTAGCACTAATATTTCTTGGCTTCAAGAACTTCAATGAGGCAATCAACATCAAA 552
Qy 492 TGCCGCAAGATTTTAAATTAATTAATGATTTCTATTAACCTTATCTGCTGTAATCT 551
Db 553 TGCAAGCAATTTTAAAGTATGATTTTAACTTATCTTATCTGCTGTAACATCT 612
Qy 552 GGAAGAGGATTTTCAAAAACTTTTACAAATGATTAAGGCTAAATCATTAAGAA 611
Db 613 GGAAGAGGATTTTCAAAAACTTTTCAAAATTTTACAAATGCTAGAACTAATCATTAAGAA 672
Qy 612 TTATATTGAGAGAAAAAACTTTCTAAACCTTTCTTTCGACACCGCAGAGATTACCTGA 671
Db 673 TTATATTGAGAGAAAAAACTTTCTAAACCTTTCTTTCGACACCGCAGAGATTACCTGA 732
Qy 672 TGGCAAGTATGATTTATTTGGCTGACCAAGAGCGCTTAA----- 714
Db 733 TGGCAAGTATGATTTATTTGGCCGGTCCAAAGCGCTTAAATTTGCCAAGAGTCTTC 792
Qy 715 -----TGGAGTGAAGTTTAAAGAACTTAA 740
Db 793 TACAGCAACAAAGCTTACAGCAGCTTCACTTAATTTGAGAGATTTGTTGCAGAAACCTTAC 852
Qy 741 AAATAACAAATCTAGAAATGATTTTCTAATATGAAAGGGCTGCAAAACAAAGTATAG 800
Db 853 AGATPACCCATCCAGAAATACATTTATGAAATATGATGATGCTGCAAAACGAAATATAG 912
Qy 801 TTCAATTATTAAGAGGTACAAAAGGTTAAGCTTCCACAGACAGACAGAGAAAGTATGG 860
Db 913 TTCAATTATTAAGAGGTACAAAAGGTTAAGTATCCAGTGCAGACAGACAGAGATTGG 972
Qy 861 TACAGCCAGTGGCAGTAACTGGAATAATGCGCAATA--ATTATATAGTGAAGGCT 917
Db 973 TACAAAAGCGGCACTTACCTTGCAGAACTGCAAGTATGAGATTTATATAGTAACT 1032
Qy 918 AAGCCAAAAGACAGGGTAACTTACTCAAAATGATATGACAAATACATGACGTTCA 977
Db 1033 AAGCCAAAAGACAGGGTAACTTCTCATTAATATATCTGACCAAAATATGAGATCCA 1092
Qy 978 TAGTGTGGAATCTAATTAATAAATATATGA 1008
Db 1093 AAGTGTGGAATCTAATTAATAAATATATGA 1123

RESULT 11

AA029373 LOCUS AA029373 1205 bp DNA linear PAT 16-SEP-2000
DEFINITION Sequence 5 from Patent WO9903328.
ACCESSION AA029373
VERSION AA029373.1 GI:10190185
KEYWORDS

SOURCE

ORGANISM

Xenorhabdus nematophila
Xenorhabdus nematophila
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Xenorhabdus.

REFERENCE

1
East, P.D.
Toxin genes from the bacteria *Xenorhabdus nematophilus* and

AUTHORS

Photorhabdus luminescens
Patent: WO 9903328-A 5 28-JAN-1999;

TITLE

COMMUNAL IND RES ORG (AU) ; EAST PETER DAVID (AU)

JOURNAL

COMMUNAL IND RES ORG (AU) ; EAST PETER DAVID (AU)

Db 678 TGGGAAAAAGATTTTTCAAAAATTTTTCAAATGCTAGTAACTAAATCATTTAGAGA 737
Qy 611 ATTATATTGAGAGAAAAAACTTTCTAAACCTTTCTTTCGACACCGAGAGATTACTG 670
Db 738 ATTATATTGAGAGAAAAAACTTTCTAAACCTTTCTTTCGACACCGAGAGATTACTG 797
Qy 671 ATGGCAGAAATAGTTATTTGGCTGGACCAAGAACGCCCTAAA----- 714
Db 798 ATGGCAGAAACAGGCTACTTGGCCGGTCCAAACAAAGCCCTAAATATGCAACAGCTCT 857
Qy 715 -----TGAGAGTGAATTTTAAAGACTT 738
Db 858 CTACAGCAACACGCTCTACAGACGCTTCTAAATTTGAGAGTTAGTTTGCACAAACCTT 917
Qy 739 AAAAATTAACAATCTAGAGATGATTTTCTAATATGAAAGGGGCTGCAAAAACAAAGTAT 798
Db 918 AGAGTAAACCATTCAGAAATACATTTTGAATAATGATGCTGCAACGAAATAT 977
Qy 799 AGTTCAATTTAATAAGAGTACAAAGGTTACGCTCCACAGACAGCAGCAAAAGTAT 858
Db 978 AGTTCAATTTAATAAGAGTACAAAGGTTATGATCCAGCTGCAAGCAGCAGCAAGTAT 1037
Qy 859 GGTACAGCCAGTGGCAGTAACTTGAAAAATTGCCAATA--ATTATATATGCTGAGG 915
Db 1038 GGTACAAAAACGGCAGTAACTTGAAAAACCTGCAAGGTAGAGATTATATATATAGA 1097
Qy 916 CTAGCCCAAAAAAGCAGGTTAACTTTACTCAAAATGATAGTAAATACATGACGTT 975
Db 1098 CTAGCCCAAAAAAGCAGGTTAACTTTCTCAATAAATTAATGACCAAAATATGAGATC 1157
Qy 976 CATAGTGTGGAACTCATTTAATAAAATATATGA 1008
Db 1158 CAAAGTGTGGAACTCATTTACCAAAATATATGA 1190

RESULT 13
AR082008 834 bp DNA linear PAT 31-AUG-2000
LOCUS AR082008
DEFINITION Sequence 3 from patent US 5972687.
ACCESSION AR082008
VERSION AR082008.1 GI:10008734
KEYWORDS
ORGANISM Unknown.
SOURCE Unknown.
REFERENCE
1 (bases 1 to 834)
AUTHORS Smigileski, A. Joseph, and Akhurst, R. Joseph.
TITLE Toxin gene from *Xenorhabdus nematophilus*
JOURNAL Patent: US 5972687-A 3 26-OCT-1999;
FEATURES
location/Qualifiers
1..834
/organism="unknown"
/mol_type="unassigned DNA"

ORIGIN
Query Match 51.7%; Score 521.4; DB 6; Length 834;
Best Local Similarity 85.4%; Pred. No. 9.4e-84;
Matches 604; Conservative 0; Mismatches 101; Indels 2; Gaps 2;

Qy 12 ATTAACACCTGATAGATGATATCCACCCGTTGAAAAAGCAATAGCAGAGATAT 71
Db 57 AGTAAACGCTGATGATTAAGTGAATATCAACCCGTTGAAAAAGCAATAGCAGAGATAT 116
Qy 72 AGTACGTACTAACTTAAGCAACAGATGAGGCTCATACGATCATATGGAATTGA 131
Db 117 AATACGTACTAGTAATTCAGCAACAAATGAAGTCAATACAGATTTGTAATGCC 176
Qy 132 ATATGAGCTAAGAAATTAATTTAGCTTAGCTTGGCTGTAAGTGTATTCATTAATG 191
Db 177 ATATGAGCTAAGAAATTAATTTAGCTTAGCTTGGCTGTAAGTGTATTCATTAATG 236
Qy 192 ATCTAACTTCTGATGACTATATTAAGAAATTAAGAGACTGCTGAGAGAAATTTACA 251

Db 237 CTCTCAACTTCAGAGACTATTTAATAAATTAAGATTAACACAGTAAATTTACAGT 296
Qy 252 ATATATGCTAATCTTTCATCTGCACTATTAGTGAATAATGATCAAAATTTCTAAAGA 311
Db 297 ATACATGCTAATCTTTCATCTGCACTATTAGTGAATAATGATCAAAATTTCTAAAGA 356
Qy 312 TATGCAAAATGCTTTTATTAAGAAATGAATGAACTGAAAGTCAATATCTCAAAACAT 371
Db 357 TATGCAAAATGCTTTTATTAAGAAATGAATGAACTGAAAGTCAATATCTCAAAACAT 415
Qy 372 TTGAATGCTTCTGAGCTTGAATAAATCAATTTGAGTCAATATCTCAAAACAT 430
Db 416 CTGGAATATCTGATCTTGAATAAATCAATTTGAGTCAATATCTCAAAACAT 475
Qy 431 TATTAAGCAATATTTTCTGTAAGAAATTTCACTGAGAAATCAATCAATCA 490
Db 476 TATTAAGCAATATTTTCTGTAAGAAATTTCACTGAGAAATCAATCAATCA 535
Qy 491 ATGCGCAGATTTTAAATTAATTAATTTCTTAATTTACCTTAATCTGCTMACTTAC 550
Db 536 ATGAGCAAAATTTTAAATTAATTAATTTCTTAATTTACCTTAATCTGCTMACTTAC 595
Qy 551 TGGGAAAGAGATTTTTCAAAAAATTTTCAATGATTAAGAGCTTAATCATTAAGA 610
Db 596 TGGGAAAGAGATTTTTCAAAAAATTTTCAATGATTAAGAGCTTAATCATTAAGA 655
Qy 611 ATTATATTGAGAGAAAAAACTTTCTAAACCTTTCTTTCGACACCGAGAGATTACTG 670
Db 656 ATTATATTGAGAGAAAAAACTTTCTAAACCTTTCTTTCGACACCGAGAGATTACTG 715
Qy 671 ATGCAAAATAGTTATTTGGCTGGAACCAAGAGGCCCTAAATGG 717
Db 716 ATGCAAAATAGTTATTTGGCTGGAACCAAGAGGCCCTAAATGG 762

RESULT 14
AR082007 837 bp DNA linear PAT 31-AUG-2000
LOCUS AR082007
DEFINITION Sequence 1 from patent US 5972687.
ACCESSION AR082007
VERSION AR082007.1 GI:10008733
KEYWORDS
ORGANISM Unknown.
SOURCE Unknown.
REFERENCE
1 (bases 1 to 837)
AUTHORS Smigileski, A. Joseph, and Akhurst, R. Joseph.
TITLE Toxin gene from *Xenorhabdus nematophilus*
JOURNAL Patent: US 5972687-A 1 26-OCT-1999;
FEATURES
location/Qualifiers
1..837
/organism="unknown"
/mol_type="unassigned DNA"

ORIGIN
Query Match 51.7%; Score 521.4; DB 6; Length 837;
Best Local Similarity 85.4%; Pred. No. 9.4e-84;
Matches 604; Conservative 0; Mismatches 101; Indels 2; Gaps 2;

Qy 12 ATTAACACCTGATAGATGATATCCACCCGTTGAAAAAGCAATAGCAGAGATAT 71
Db 57 AGTAAACGCTGATGATTAAGTGAATATCAACCCGTTGAAAAAGCAATAGCAGAGATAT 116
Qy 72 AGTACGTACTAACTTAAGCAACAGATGAGGCTCATACGATCATATGGAATTGA 131
Db 117 AATACGTACTAGTAATTCAGCAACAAATGAAGTCAATACAGATTTGTAATGCC 176
Qy 132 ATATGAGCTAAGAAATTAATTTAGCTTAGCTTGGCTGTAAGTGTATTCATTAATG 191
Db 177 ATATGAGCTAAGAAATTAATTTAGCTTAGCTTGGCTGTAAGTGTATTCATTAATG 236
Qy 192 ATCTAACTTCTGATGACTATATTAAGAAATTAAGAGACTGCTGAGAGAAATTTACA 251

